

10/820,964

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 4 年    2 月 1 0 日  
Date of Application:

出 願 番 号                      特 願 2 0 0 4 - 0 3 2 8 1 0  
Application Number:

ST. 10/C] :                      [ J P 2 0 0 4 - 0 3 2 8 1 0 ]

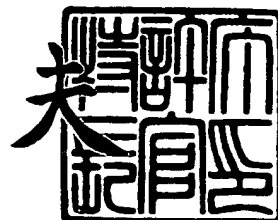
願                      人                      株式会社日立製作所  
Applicant(s):

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2 0 0 4 年    3 月    9 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号    出証特 2 0 0 4 - 3 0 1 8 0 8 6

【書類名】 特許願  
【整理番号】 K04000081A  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 12/00  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 藤本 和久  
【発明者】  
    【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
                                D システム事業部内  
    【氏名】 井上 靖雄  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 細谷 睦  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 島田 健太郎  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                                システム開発研究所内  
    【氏名】 渡▲辺▼ 直企  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社 日立製作所  
【代理人】  
    【識別番号】 100075096  
    【弁理士】  
    【氏名又は名称】 作田 康夫  
【選任した代理人】  
    【識別番号】 100100310  
    【弁理士】  
    【氏名又は名称】 井上 学  
【手数料の表示】  
    【予納台帳番号】 013088  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

**【書類名】 特許請求の範囲****【請求項 1】**

計算機またはディスク装置と接続される接続部を有するインターフェース部と、メモリ部と、プロセッサ部と、ディスク装置とを有し、前記インターフェース部、前記メモリ部及び前記プロセッサ部の間は相互結合網で互いに接続されていることを特徴とするストレージシステム。

**【請求項 2】**

前記メモリ部は、前記計算機または前記ディスク装置との間で読み出し又は書き込まれるデータを格納するキャッシュメモリ及び制御情報を格納する制御メモリを有し、

前記プロセッサ部は、前記計算機と前記ディスク装置との間でのデータの該ストレージシステム内での転送を制御する複数のマイクロプロセッサを有することを特徴とする請求項 1 記載のストレージシステム。

**【請求項 3】**

前記複数のマイクロプロセッサは、該ストレージシステム内でのデータ転送を制御する際、制御対象となる前記インターフェース部又は前記メモリ部へ、前記相互結合網を介して前記制御情報を転送することを特徴とする請求項 2 載のストレージシステム。

**【請求項 4】**

前記相互結合網が、データを転送する相互結合網と前記制御情報を転送する相互結合網とを有することを特徴とする請求項 3 記載のストレージシステム。

**【請求項 5】**

前記相互結合網が複数のスイッチ部を有することを特徴とする請求項 4 記載のストレージシステム。

**【請求項 6】**

前記複数のマイクロプロセッサのうちの何れかは、前記インターフェース部と前記メモリ部との間のデータ転送の制御のみを実行することを特徴とする請求項 5 記載のストレージシステム。

**【請求項 7】**

前記複数のマイクロプロセッサのうちの第一のマイクロプロセッサは、前記計算機に接続されるインターフェース部と前記メモリ部との間のデータ転送の制御のみを実行し、前記複数のマイクロプロセッサのうちの第二のマイクロプロセッサは、前記ディスク装置に接続されるインターフェース部と前記メモリ部との間のデータ転送の制御のみを実行することを特徴とする請求項 6 記載のストレージシステム。

**【請求項 8】**

複数のクラスタを有し、前記複数のクラスタの各々は、計算機またはディスク装置との接続部を有するインターフェース部と、前記計算機または前記ディスク装置との間で送受信されるデータを格納するキャッシュメモリ及び制御情報を格納する制御メモリとを有するメモリ部と、前記計算機と前記ディスク装置との間でのデータの転送を制御するマイクロプロセッサを有するプロセッサ部と、ディスク装置とを有し、前記複数のクラスタの個々が有する前記インターフェース部、前記メモリ部及び前記プロセッサ部は、相互結合網により前記複数のクラスタのうち他のクラスタが有する前記インターフェース部、前記メモリ部及び前記プロセッサ部に接続されることを特徴とするストレージシステム。

**【請求項 9】**

前記複数のクラスタの個々はスイッチ部を有し、前記複数のクラスタの個々が有する前記インターフェース部、前記メモリ部及び前記プ

ロセッサ部は、前記スイッチ部を用いて前記クラスタ内で相互に接続され、

前記複数のクラスタは、前記スイッチ部間を接続することにより相互に接続されていることを特徴とする請求項 8 記載のストレージシステム。

【請求項 10】

前記スイッチ部間を他のスイッチを用いて接続することを特徴とする請求項 9 記載のストレージシステム。

【請求項 11】

前記計算機が要求するデータは、前記複数のクラスタのうち前記計算機が接続される第一のクラスタとは異なる第二のクラスタが有するディスク装置に格納されていることを特徴とする請求項 10 記載のストレージシステム。

【請求項 12】

前記計算機が要求するデータが、前記複数のクラスタのうち前記計算機が接続される第一のクラスタとは異なる第二のクラスタが有するディスク装置に格納されている場合、前記第一のクラスタの前記プロセッサ部は、前記第二のクラスタの前記インターフェース部に対して、前記スイッチ部を介してデータ転送の指示を送信することを特徴とする請求項 11 記載のストレージシステム。

【請求項 13】

前記インターフェース部が第 1 の回路基板に実装され、

前記メモリ部が第 2 の回路基板に実装され、

前記プロセッサ部が第 3 の回路基板に実装され、

前記スイッチ部が第 4 の回路基板に実装され、

更に、前記第 1、第 2、第 3 及び第 4 の回路基板間を接続する信号線がプリントされ、プリントされた前記信号線に前記第 1、第 2、第 3 及び第 4 の回路基板を接続するための第 1 のコネクタを有する 1 つのバックプレーンを有し、

前記第 1、第 2、第 3 及び第 4 の回路基板は前記バックプレーンの前記第 1 のコネクタに接続するための第 2 のコネクタを有することを特徴とする請求項 5 記載のストレージシステム。

【請求項 14】

前記バックプレーンに接続できる前記回路基板の総数が  $n$  で、前記第 4 の回路基板の数と接続場所は予め定められており、前記第 1、第 2、第 3 及び第 4 の回路基板の総数が  $n$  を超えない範囲で前記バックプレーンに接続する前記第 1、第 2、及び第 3 の回路基板それぞれの数を自由に選択できることを特徴とする請求項 13 記載のストレージシステム。

【請求項 15】

前記クラスタの各々は、

前記インターフェース部が実装される第 1 の回路基板と、

前記メモリ部が実装される第 2 の回路基板と、

前記プロセッサ部が実装される第 3 の回路基板と、

前記スイッチ部が実装される第 4 の回路基板と、

前記第 1、第 2、第 3 及び第 4 の回路基板間を接続する信号線がプリントされ、プリントされた前記信号線に前記第 1、第 2、第 3 及び第 4 の回路基板を接続するための第 1 のコネクタを有する 1 つのバックプレーンとを有し、

前記第 1、第 2、第 3 及び第 4 の回路基板は前記バックプレーンの前記第 1 のコネクタに接続するための第 2 のコネクタを有することを特徴とする請求項 9 記載のストレージシステム。

【請求項 16】

前記複数のクラスタの数と前記バックプレーンの数が等しいことを特徴とする請求項 15 記載のストレージシステム。

【請求項 17】

前記第 4 の回路基板はケーブルを接続するための第 3 のコネクタを有し、かつ前記第 3 のコネクタと前記スイッチ部を接続する信号線が基板上に配線されており、

前記複数のクラスタ間は、前記第3のコネクタ間を前記ケーブルによって接続することで相互に接続されることを特徴とする請求項16記載のストレージシステム。

【請求項18】

前記インターフェース部は第1の回路基板上に実装され、  
前記メモリ部、前記プロセッサ部及び前記スイッチ部は第5の回路基板上に実装され、  
更に前記第1と前記第5の回路基板間を接続する信号線がプリントされ、プリントされた前記信号線に前記第1と前記第5の回路基板を接続するための第4のコネクタを有するバックプレーンを有し、

前記第1と前記第5の回路基板は前記バックプレーンの前記第4のコネクタに接続するための第5のコネクタを有することを特徴とする請求項5記載のストレージシステム。

【請求項19】

前記インターフェース部、前記メモリ部、前記プロセッサ部及び前記スイッチ部は、第6の回路基板上に実装されていることを特徴とする請求項5記載のストレージシステム。

【請求項20】

計算機またはディスク装置と接続される接続部を有するインターフェース部と、  
メモリ部と、  
プロセッサ部と、  
ディスク装置とを有し、

前記インターフェース部、前記メモリ部及び前記プロセッサ部の間は相互結合網で互いに接続されており、

前記計算機からデータ読み出しのコマンドを受信した前記インターフェースは、前記受信したコマンドを前記プロセッサ部へ転送し、

前記プロセッサ部は、前記コマンドを解析して前記コマンドで要求されるデータの格納場所を特定し、前記メモリ部へアクセスして前記コマンドで要求されるデータが前記メモリ部に格納されているかどうかを確認し、

前記メモリ部に前記コマンドで要求されるデータが格納されている場合、前記プロセッサ部は、前記インターフェース部に前記メモリ部から前記要求されるデータの読み出しを前記相互結合網を介して指示し、

前記インターフェース部は前記プロセッサ部の指示に従って、前記要求されるデータを前記メモリ部から前記相互結合網を介して読み出して前記計算機へ転送し、

前記メモリ部に前記コマンドで要求されるデータが格納されていない場合、前記プロセッサ部は、前記要求されるデータが格納される前記ディスク装置が接続される前記インターフェース部に前記ディスク装置から前記要求されたデータを読み出して前記メモリ部へ格納するよう、前記相互結合網を介して指示し、

前記ディスク装置が接続される前記インターフェース部は、前記プロセッサ部からの指示に基づいて、前記要求されたデータを前記ディスク装置から読み出して前記相互結合網を介して前記メモリ部に転送し、転送の終了を前記プロセッサ部に送信し、

前記プロセッサ部は、前記転送の終了を受信後、前記計算機が接続される前記インターフェース部に前記メモリ部から前記要求されたデータを読み出して前記計算機へ転送することを前記相互結合網を介して指示し、

前記計算機が接続される前記インターフェース部は、前記プロセッサ部の指示に基づいて、前記メモリ部から前記要求されたデータを前記相互結合網を介して読み出して前記計算機へ転送することを特徴とするストレージシステム。

**【書類名】明細書****【発明の名称】**ストレージシステム**【技術分野】****【0001】**

本発明は、小規模から大規模までスケラブルに構成を拡張可能なストレージシステムに関する。

**【背景技術】****【0002】**

昨今、情報処理システムで処理されるデータを保存するストレージシステムが、情報処理システムの中心的な役割を担うようになっている。ストレージシステムには、小規模な構成から大規模な構成まで数多くの種類のシステムがある。

**【0003】**

例えば、図20に示すような構成のストレージシステムが特許文献1に開示されている。このストレージシステムは、計算機（以下「サーバ」とも言う）3との間のデータ転送を実行する複数のチャンネルインターフェース（以下「IF」とも言う）部11、ハードディスク群2との間のデータ転送を実行する複数のディスクIF部16、ハードディスク群2に格納するデータを一時的に格納するキャッシュメモリ部14、ストレージシステム8に関する制御情報（例えば、ストレージシステム8内のデータ転送制御に関する情報、ハードディスク群2に格納するデータの管理情報等）を格納する制御メモリ部15及びハードディスク群2とを有する。そして、チャンネルIF部11、ディスクIF部16及びキャッシュメモリ部14との間は相互結合網41で接続され、チャンネルIF部11、ディスクIF部16及び制御メモリ部15との間は相互結合網42で接続されている。また、相互結合網41や相互結合網42は共通バスやスイッチで構成される。

**【0004】**

特許文献1記載のストレージシステムでは、上述の構成により、1つのストレージシステム8内において、キャッシュメモリ部14および制御メモリ部15は全てのチャンネルIF部11及びディスクIF部16からアクセス可能な構成となっていた。

**【0005】**

また、特許文献2に開示されている従来技術では、図21に示すように、複数のディスクアレイ装置4がディスクアレイスイッチ5を介して複数のサーバ3に接続され、ディスクアレイスイッチ5及び各ディスクアレイ装置4に接続されたシステム構成管理手段60で、複数のディスクアレイ装置4を1つのストレージシステム9として管理する。

**【0006】****【特許文献1】** 米国特許第6385681号**【0007】****【特許文献2】** 米国特許第6542961号**【発明の開示】****【発明が解決しようとする課題】****【0008】**

企業は情報処理システムへの初期投資を抑え、ビジネス規模の拡張に応じて情報処理システムの拡張を行う傾向にある。このため、初期投資が小さくかつ事業規模に合わせてリーズナブルな投資で規模を拡張するためのコストおよび性能のスケラビリティがストレージシステムに要求される。ここで従来技術の性能のスケラビリティ及びコストについて検討する。

**【0009】**

ストレージシステムに要求される性能（単位時間当たりのデータの入出力回数や単位時間当たりのデータの転送量）は年々向上している。従って、将来的な性能向上に対応するため、特許文献1のストレージシステムが有するチャンネルIF部11およびディスクIF部16のデータ転送処理性能も向上させる必要がある。

**【0010】**

しかし特許文献 1 の技術では、全てのチャネル I F 部 1 1 と全てのディスク I F 部 1 6 が、キャッシュメモリ部 1 4 および制御メモリ部 1 5 を介して、チャネル I F 部 1 1 とディスク I F 部 1 6 の間のデータ転送を制御する。従ってチャネル I F 部 1 1 およびディスク I F 部 1 6 のデータ転送処理性能が向上すると、キャッシュメモリ部 1 4 や制御メモリ部へのアクセス負荷が増大する。するとこのアクセス負荷がネックとなり、ストレージシステム 8 の性能を向上することが将来的に難しくなる、すなわち性能のスケーラビリティを確保できなくなる。

#### 【0011】

一方特許文献 2 の技術では、ディスクアレイスイッチ 5 のポート数を増やすことや複数のディスクアレイスイッチ 5 を多段に接続することで、接続可能なディスクアレイ装置 4 およびサーバ 3 の数を増やすことができる。すなわち性能のスケーラビリティを確保することができる。

#### 【0012】

しかしながら、特許文献 2 の技術では、サーバ 3 はディスクアレイスイッチ 5 を介してディスクアレイ装置 4 にアクセスする。したがって、ディスクアレイスイッチ 5 が有するサーバ 3 とのインターフェース部においてサーバとディスクアレイスイッチとの間のプロトコルがディスクアレイスイッチ内のプロトコルに変換され、さらにディスクアレイスイッチ 5 が有するディスクアレイ装置 4 とのインターフェース部においてディスクアレイスイッチ内のプロトコルがディスクアレイスイッチとディスクアレイ装置との間のプロトコルに変換されるという 2 回のプロトコル変換処理が発生する。従ってディスクアレイスイッチを介さずに直接ディスクアレイ装置にアクセス可能な場合に比べ、応答性能が劣る。

#### 【0013】

またコストを考慮しなければ、特許文献 1 において、キャッシュメモリ部 1 4 や制御メモリ部を大規模化して許容可能なアクセス性能を向上させることは可能である。しかし、全てのチャネル I F 部 1 1 およびディスク I F 部 1 6 からキャッシュメモリ部 1 4 や制御メモリ部 1 5 をアクセス可能とするためにキャッシュメモリ部 1 4 や制御メモリ部 1 5 をそれぞれ 1 つの共有メモリ空間として管理する必要がある。このため、キャッシュメモリ部 1 4 や制御メモリ部 1 5 を大規模化すると、小規模構成におけるストレージシステムの低コスト化が難しく、小規模構成のストレージシステムを低価格で提供することが難しい。

#### 【課題を解決するための手段】

#### 【0014】

上述した課題を解決するために、本発明の一実施態様は以下の構成を有する。具体的には、計算機またはディスク装置との接続部を有するインターフェース部、計算機またはディスク装置との間で送受信されるデータや制御情報を格納するメモリ部、計算機とディスク装置との間でデータの転送を制御するマイクロプロセッサを有するプロセッサ部、及びディスク部を有するストレージシステムであって、インターフェース部、メモリ部、プロセッサ部の間が相互結合網で互いに接続されている構成である。

#### 【0015】

そして、本発明のストレージシステムでは、プロセッサ部がインターフェース部とメモリ部との間で制御情報をやり取りすることにより、プロセッサ部が、計算機から要求されたデータ読み出し又はデータの書き込みにかかるデータの転送を指示する。

#### 【0016】

尚、相互結合網の一部または全部を、データを転送する相互結合網と制御情報を転送する相互結合網に分離する構成としても良い。又、相互結合網は複数のスイッチ部から構成されていても良い。

#### 【0017】

また本発明の他の実施態様として、以下の構成がある。具体的には、複数のクラスタが通信網を介して接続されるストレージシステムである。ここで個々のクラスタは、計算機またはディスク装置との接続部を有するインターフェース部、計算機またはディスク装置

との間でリード／ライトされるデータやシステムの制御情報を格納するメモリ部、計算機とディスク装置との間でのデータのリード／ライトを制御するマイクロプロセッサを有するプロセッサ部、及びディスク部を有する。そして、各クラスタ内のインターフェース部、メモリ部及びプロセッサ部は、通信網を介して他のクラスタ内の各部と接続される構成とする。

【0018】

尚、各クラスタ内のインターフェース部、メモリ部及びプロセッサ部は、少なくとも1つのスイッチ部によりクラスタ内で接続され、接続パスにより各クラスタのスイッチ部間を相互に接続する構成としても良い。

【0019】

更に、各クラスタが有するスイッチ部間を別のスイッチを介して接続することで、各クラスタ間を接続する構成としても良い。

【0020】

又、他の実施態様として、上述した実施態様におけるインターフェース部がプロトコル処理用のプロセッサを有する構成としても良い。この場合、インターフェース部においてプロトコル処理を行い、プロセッサ部においてストレージシステム内におけるデータ転送を制御する構成としても良い。

【0021】

その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【発明の効果】

【0022】

本発明によれば、サーバ接続数やハードディスク接続数やシステム性能に対するユーザの要求に柔軟に対応可能なフレキシブルな構成のストレージシステムを提供することが可能となる。また、ストレージシステムの共有メモリネックを解消するとともに、小規模な構成の低コスト化を図り、小規模から大規模な構成までコストと性能のスケラビリティを実現可能なストレージシステムを提供することが可能となる。

【発明を実施するための最良の形態】

【0023】

以下、本発明の実施形態を図面を用いて説明する。

【実施例1】

【0024】

図1は、第一の実施形態のストレージシステムの構成例を示す図である。ストレージシステム1は、サーバ3あるいはハードディスク群2とのデータの送受信を行うインターフェース部10、プロセッサ部81、メモリ部21及びハードディスク群2を有する。インターフェース部10、プロセッサ部81及びメモリ部21の間は相互結合網31を介して接続されている。

相互結合網31の具体的な構成の一例を図2に示す。

【0025】

相互結合網31は2つのスイッチ部51を有する。インターフェース部10、プロセッサ部81及びメモリ部21は、2つのスイッチ部51と1本ずつの通信パスで各々接続される。ここで、通信パスとはデータや制御情報を伝送するための1つまたは複数の信号線から成る伝送路である。これにより、インターフェース部10、プロセッサ部81及びメモリ部21のそれぞれの間で2つの通信経路を確保し、信頼性をあげることが可能となる。なお、ここで、上記個数や本数は一実施例に過ぎず、個数を上記に限定するものではない。このことは以下に説明する実施形態全てに当てはまる。

また、相互結合網はスイッチを利用したものを例に示したが、相互に接続され制御情報やデータが転送されれば良いのであり、例えばバスで構成されても良い。

【0026】

また、図3に示すように、相互結合網31を、データを転送する相互結合網41と制御



情報を転送する相互結合網 4 2 に分離しても良い。こうすることにより、データと制御情報を 1 つの通信パスで転送する場合(図 1)に比べ、データと制御情報の転送が相互に干渉することが無くなる。これによってデータ及び制御情報の転送性能を向上することが可能となる。

#### 【0027】

図 4 は、相互結合網 4 1、4 2 の具体的な構成の一例を示す図である。相互結合網 4 1、4 2 は、それぞれ 2 つのスイッチ部 5 2、5 6 を有する。インターフェース部 1 0、プロセッサ部 8 1 及びメモリ部 2 1 は、2 つのスイッチ部 5 2 及び 2 つのスイッチ部 5 6 の各々と 1 本ずつの通信パスで接続される。これにより、インターフェース部 1 0、プロセッサ部 8 1、およびメモリ部 2 1 のそれぞれの間でデータ用パス 9 1 を 2 本、制御情報用パス 9 2 を 2 本それぞれ確保し、信頼性をあげることが可能となる。

#### 【0028】

図 8 は、インターフェース部 1 0 の構成の具体例を示す図である。

インターフェース部 1 0 は、サーバ 3 あるいはハードディスク群 2 と接続される 4 つの I F (外部 I F) 1 0 0、プロセッサ部 8 1 あるいはメモリ部 2 1 との間のデータ／制御情報の転送を制御する転送制御部 1 0 5 及びデータのバッファリングや制御情報の格納を行うメモリモジュール 1 2 3 を有する。

#### 【0029】

外部 I F 1 0 0 は転送制御部 1 0 5 と接続される。また、メモリモジュール 1 2 3 は転送制御部 1 0 5 に接続される。転送制御部 1 0 5 は、メモリモジュール 1 2 3 へのデータ／制御情報のリード／ライトを制御するメモリコントローラとしても動作する。

#### 【0030】

ここで、外部 I F 1 0 0 あるいはメモリモジュール 1 2 3 と転送制御部 1 0 5 の間の接続構成は一実施例に過ぎず、構成を上記に限定するものではない。少なくとも、外部 I F 1 0 0 から転送制御部 1 0 5 を経てプロセッサ部 8 1、メモリ部 2 1 へデータ／制御情報を転送可能な構成であれば良い。

#### 【0031】

なお、図 4 に示すデータ用パス 9 1 と制御情報用パス 9 2 を分離した場合のインターフェース部 1 0 においては、転送制御部 1 0 5 にはデータ用パス 9 1 が 2 本、制御情報用パス 9 2 が 2 本接続される。

#### 【0032】

図 9 は、プロセッサ部 8 1 の構成の具体例を示す図である。

プロセッサ部 2 1 は、2 つのマイクロプロセッサ 1 0 1、インターフェース部 1 0 あるいはメモリ部 2 1 との間のデータ／制御情報の転送を制御する転送制御部 1 0 5 及びメモリモジュール 1 2 3 を有する。メモリモジュール 1 2 3 は転送制御部 1 0 5 に接続される。転送制御部 1 0 5 は、メモリモジュール 1 2 3 へのデータ／制御情報のリード／ライトを制御するメモリコントローラとしても動作する。メモリモジュール 1 2 3 は、2 つのマイクロプロセッサ 1 0 1 の主記憶として共有され、データや制御情報を格納する。なお、プロセッサ部 2 1 は、2 つのマイクロプロセッサ 1 0 1 に共有されるメモリモジュール 1 2 3 の代わりに、各マイクロプロセッサ 1 0 1 専用のメモリモジュールをマイクロプロセッサの数だけ有しても良い。

#### 【0033】

マイクロプロセッサ 1 0 1 は転送制御部 1 0 5 に接続される。マイクロプロセッサ 1 0 1 は、メモリ部 2 1 の制御メモリモジュール 1 2 7 内に格納された制御情報に基づいて、メモリ部 2 1 が有するキャッシュメモリへのデータのリード／ライト、キャッシュメモリのディレトリ管理、インターフェース部 1 0 とメモリ部 2 1 との間のデータ転送を制御する。

具体的には、例えばインターフェース部 1 0 内の外部 I F 1 0 0 が、データのリードまたはライトのアクセス要求を示す制御情報をプロセッサ部 8 1 内のメモリモジュール 1 2 3 に書き込む。その後、マイクロプロセッサ 1 0 1 は書き込まれた制御情報を読み出して

解釈し、外部 I F 1 0 0 からどのメモリ部 2 1 へデータを転送するかを示す制御情報及びそのデータ転送に必要なパラメータをインターフェース部 1 0 内のメモリモジュール 1 2 3 へ書き込む。外部 I F 1 0 0 はその制御情報とパラメータに従い、メモリ部 2 1 へのデータ転送を実行する。

#### 【0034】

またマイクロプロセッサ 1 0 1 は、インターフェース部 1 0 に接続されたハードディスク群 2 へ書き込むデータの冗長化処理、いわゆる R A I D 処理を実行する。この R A I D 処理は、インターフェース部 1 0 やメモリ部 2 1 において実行しても問題ない。更にマイクロプロセッサ 1 0 1 は、ストレージシステム 1 における記憶領域の管理（論物変換等）も行う。

#### 【0035】

ここで、マイクロプロセッサ 1 0 1、転送制御部 1 0 5 及びメモリモジュール 1 2 3 の間の接続構成は一例に過ぎず、構成を上記に限定するものではない。少なくともマイクロプロセッサ 1 0 1、転送制御部 1 0 5 及びメモリモジュール 1 2 3 の間で相互にデータを転送できる構成であれば良い。

なお、図 4 に示すようにデータ用バス 9 1 と制御情報用バス 9 2 を分離した場合、プロセッサ部 8 1 の転送制御部 1 0 6 にはデータ用バス 9 1（ここでは 2 本）と制御情報用バス 9 2（ここでは 2 本）とが接続される。

#### 【0036】

図 1 0 は、メモリ部 2 1 の構成の具体的例を示す図である。

メモリ部 2 1 は、キャッシュメモリモジュール 1 2 6、制御メモリモジュール 1 2 7 及びメモリコントローラ 1 2 5 を有する。キャッシュメモリモジュール 1 2 6 には、ハードディスク群 2 に書き込むデータまたはハードディスク群 2 から読み出したデータが一時的に格納（以下「キャッシング」）される。制御メモリモジュール 1 2 7 には、キャッシュメモリモジュール 1 2 6 のディレクトリ情報（キャッシュメモリ上のデータを格納する論理的な区画に関する情報）、インターフェース部 1 0、プロセッサ部 8 1 及びメモリ部 2 1 間のデータ転送を制御するための情報、ストレージシステム 1 の管理情報並びに構成情報等が格納される。メモリコントローラ 1 2 5 は、キャッシュメモリモジュール 1 2 6 と制御メモリモジュール 1 2 7 へのデータのリード／ライトの処理を独立に制御する。

また、メモリコントローラ 1 2 5 は、インターフェース部 1 0、プロセッサ部 8 1 及び他のメモリ部 2 1 との間のデータ／制御情報の転送を制御する。

#### 【0037】

ここで、キャッシュメモリモジュール 1 2 6 と制御メモリモジュール 1 2 7 を物理的に 1 つにまとめ、1 つのメモリ空間上の論理的に異なる領域にキャッシュメモリ領域と制御メモリ領域を割り当てても良い。こうすることにより、メモリモジュール数を減らすことができ、部品コストを削減することが可能となる。

#### 【0038】

また、メモリコントローラ 1 2 5 をキャッシュメモリモジュール制御用と制御メモリモジュール制御用に分離しても良い。

#### 【0039】

ここで、ストレージシステム 1 がメモリ部 2 1 を複数有する場合、複数のメモリ部 2 1 を 2 つのグループに分け、そのグループ間でキャッシュメモリモジュール及び制御メモリモジュールへ格納するデータや制御情報を二重化しても良い。こうすることにより、1 つのグループのキャッシュメモリモジュールまたは制御メモリモジュールに障害が発生した場合、もう一方のグループのキャッシュメモリモジュールまたは制御メモリモジュールに格納されたデータ等で動作を継続することが可能となり、ストレージシステム 1 の信頼性が向上する。

#### 【0040】

なお、図 4 に示すようにデータ用バス 9 1 と制御情報用バス 9 2 を分離した場合、メモリコントローラ 1 2 5 にはデータ用バス 9 1（ここでは 2 本）と制御情報用バス 9 2（こ

こでは2本)とが接続される。

#### 【0041】

図11は、スイッチ部51の構成の具体例を示す図である。

スイッチ部51はスイッチLSI58を有する。スイッチLSI58は、4つのパスIF130、ヘッダ解析部131、アービタ132、クロスバスイッチ133、8つのバッファ134及び4つのパスIF135を有する。

#### 【0042】

パスIF130は、インターフェース部10と接続される通信パスが接続されるIFである。インターフェース部10及びパスIF130は、一対一に接続される。パスIF135は、プロセッサ部81又はメモリ部21と接続される通信パスが接続されるIFである。プロセッサ部81又はメモリ部21とパスIF135は、一対一に接続される。バッファ134には、インターフェース部10、プロセッサ部81及びメモリ部21の間で転送されるパケットが一時的に格納(バッファリング)される。

#### 【0043】

図12は、インターフェース部10、プロセッサ部81及びメモリ部21の間で転送されるパケットのフォーマットの一例を示す図である。パケットとは、各部の間でのデータ(制御情報を含む)転送の際に使用されるプロトコルにおけるデータ転送の単位である。パケット200は、ヘッダ210、ペイロード220及びエラーチェックコード230を有する。ヘッダ210には、少なくともパケットの送信元と送信先を示す情報が格納される。ペイロード220には、コマンド、アドレス、データ、ステータス等の情報が格納される。また、エラーチェックコード230は、パケット転送時にパケット内で発生するエラーの検出のために使用するコードである。

#### 【0044】

スイッチLSI58は、パスIF130又は135がパケットを受信すると、受信したパケットのヘッダ210をヘッダ解析部131へ送る。ヘッダ解析部131は、ヘッダ210に含まれるパケットの送信先の情報に基づいて、各パスIF間の接続要求を割り出す。具体的には、ヘッダ解析部131は、ヘッダ210で指定されるパケット送信先の装置(メモリ部等)と接続されているパスIFを割り出し、パケットを受信したパスIFと割り出されたパスIFとの間での接続要求を生成する。

#### 【0045】

その後、ヘッダ解析部131は、生成した接続要求をアービタ132へ送る。アービタ132は、割り出された各パスIFの接続要求を元に各パスIF間の調停(アービトレーション)を行う。その結果に基づいて、アービタ132は、クロスバスイッチ133に対して接続切換を示す信号を出力する。信号を受信したクロスバスイッチ133は、信号の内容に基づいてクロスバスイッチ133内の接続を切り換え、所望のパスIF間の接続を実現する。

#### 【0046】

ここで、本実施例では各パスIFに一対一にバッファを持つ構成としたが、スイッチLSI58が1つの大きなバッファを持ち、その中から各パスIFにパケット格納領域を割り当てるという構成にしても良い。また、スイッチLSI58はスイッチ部51内の障害情報を格納するメモリを有する。

#### 【0047】

図16は、相互結合網31の他の構成例を示す図である。

図16では、スイッチ部51のパスIFの数を10に増やし、かつスイッチ部51の数を4に増やした。その結果、インターフェース部10、プロセッサ部81及びメモリ部21の数が図2の構成の倍になる。尚、図16においては、インターフェース部10は一部のスイッチ部51にしか接続されないが、プロセッサ部81及びメモリ部21は全てのスイッチ部51に接続される構成である。こうしても、全てのインターフェース部10から全てのメモリ部21及び全てのプロセッサ部81へのアクセスが可能となる。

#### 【0048】

また逆に、インターフェース部 10 個々が全てのスイッチ部 51 と接続され、プロセッサ部 81 及びメモリ部 21 の各々は一部のスイッチ部 51 と接続される構成としても良い。例えば、プロセッサ部 81 及びメモリ部 21 を 2 つのグループに分け、1 つのグループは 2 つのスイッチ部 51 と接続され、他のグループは残りの 2 つのスイッチ部 51 と接続される構成とする。こうすることによっても、全てのインターフェース部 10 から全てのメモリ部 21 及び全てのプロセッサ部 81 へのアクセスが可能となる。

#### 【0049】

次に、サーバ 3 からストレージシステム 1 のハードディスク群 2 に記録されたデータを読み出す場合の処理手順の例を述べる。なお、以下の説明においてスイッチ 51 を使用するデータ転送には全てパケットが使用される。又、プロセッサ部 81 とインタフェース部 10 との通信において、プロセッサ部 81 から送信された制御情報（データ転送等に必要な情報）をインターフェース部 10 が格納する場所は、あらかじめ決まっている。

#### 【0050】

図 22 は、サーバ 3 からストレージシステム 1 のハードディスク群 2 に記録されたデータを読み出す場合の処理手順例を示したフローチャートである。

まずサーバ 3 は、ストレージシステム 1 に対してデータの読出しコマンドを発行する。インターフェース部 10 内の外部 I/F 100 がコマンドを受信する（742）と、コマンド待ち（741）にあった外部 I/F 100 は、受信したコマンドを転送制御部 105 及び相互結合網 31（ここではスイッチ部 51 とする）を介してプロセッサ部 81 内の転送制御部 105 へ転送する。コマンドを受信した転送制御部 105 は、メモリモジュール 123 に受信したコマンドを書き込む。

#### 【0051】

プロセッサ部 81 のマイクロプロセッサ 101 は、メモリモジュール 123 へコマンドが書き込まれたことをメモリモジュール 123 へのポーリング、または転送制御部 105 からの書き込みを示す割り込みによって検出する。コマンドの書き込みを検知したマイクロプロセッサ 101 は、メモリモジュール 123 から該当するコマンドを読み出してコマンド解析を行う（743）。マイクロプロセッサ 101 は、コマンド解析の結果、サーバ 3 が要求するデータが記録されている記憶領域を示す情報を割り出す（744）。

#### 【0052】

マイクロプロセッサ 101 は、コマンド解析によって得られた記憶領域の情報及びプロセッサ部 81 内のメモリモジュール 123 又はメモリ部 21 内の制御メモリモジュール 127 に格納されているキャッシュメモリモジュールのディレクトリ情報から、メモリ部 21 内のキャッシュメモリモジュール 126 に、コマンドで要求されるデータ（以下「要求データ」とも言う）が記録されているかどうかを確認する（745）。

#### 【0053】

キャッシュメモリモジュール 126 に要求データがあった場合（以下「キャッシュヒット」とも言う。）（746）、マイクロプロセッサ 101 は、キャッシュメモリモジュール 126 からインターフェース部 10 内の外部 I/F 100 へ要求データを転送するために必要な情報、具体的には要求データが格納されているキャッシュメモリモジュール 126 内のアドレス及び転送先となるインターフェース部 10 が有するメモリモジュール 123 内のアドレスの情報を、プロセッサ部 81 内の転送制御部 105、スイッチ部 51 及びインターフェース部 10 内の転送制御部 105 を介してインターフェース部 10 内のメモリモジュール 123 へ転送する。

#### 【0054】

その後、マイクロプロセッサ 101 は、外部 I/F 100 にメモリ部 21 からデータを読み出すように指示する（752）。

#### 【0055】

指示を受けたインターフェース部 10 内の外部 I/F 100 は、まず自インターフェース部 10 内のメモリモジュール 123 の所定の場所から要求データの転送に必要な情報を読み出す。その情報をもとに、インターフェース部 10 内の外部 I/F 100 はメモリ部 21

のメモリコントローラ125へアクセスして、キャッシュメモリモジュール126からの要求データの読み出しを要求する。要求を受けたメモリコントローラ125は、キャッシュメモリモジュール126から要求データを読み出し、その要求データを、要求を受けたインターフェース部10へ転送する(753)。要求データを受信したインターフェース部10は、受信した要求データをサーバ3へ送出する(754)。

#### 【0056】

一方、キャッシュメモリモジュール126に要求データがない場合(以下「キャッシュミス」とも言う。)(746)、まずマイクロプロセッサ101は、メモリ部21内の制御メモリモジュール127へアクセスし、キャッシュメモリモジュールのディレクトリ情報に、メモリ部21内のキャッシュメモリモジュール126に要求データを格納する領域を確保するための情報、具体的には空いているキャッシュスロットを指定する情報を登録する(以下「キャッシュ領域確保」とも言う。)(747)。キャッシュ領域確保後、マイクロプロセッサ101は、メモリ部21内の制御メモリモジュール127へアクセスし、制御メモリモジュール127に格納されている記憶領域の管理情報から、要求データが格納されているハードディスク群2が接続されているインターフェース部10(以下「目的インターフェース部10」とも言う)を割り出す(748)。

#### 【0057】

その後、マイクロプロセッサ101は、目的インターフェース部10内の外部IF100からキャッシュメモリモジュール126へ要求データを転送するために必要な情報を、プロセッサ部81内の転送制御部105、スイッチ部51及び目的インターフェース部10内の転送制御部105を介して、目的インターフェース部10内のメモリモジュール123へ転送する。そしてマイクロプロセッサ101は、ハードディスク群2から要求データを読み出しメモリ部21へ要求データを書き込むよう、目的インターフェース部10内の外部IF100へ指示する。

#### 【0058】

指示を受けた目的インターフェース部10内の外部IF100は、指示に基づいて自インターフェース部10内のメモリモジュール123の所定の場所から要求データの転送に必要な情報を読み出す。その情報をもとに、目的インターフェース部10内の外部IF100はハードディスク群2から要求データを読み出し(749)、読み出したデータをメモリ部21内のメモリコントローラ125へ転送する。メモリコントローラ125は、受信した要求データをキャッシュメモリモジュール126へ書き込む(750)。要求データの書き込みが終了すると、メモリコントローラ125は、その終了をプロセッサ101へ通知する。

#### 【0059】

キャッシュメモリモジュール126への書き込みの終了を検出したマイクロプロセッサ101は、メモリ部21内の制御メモリモジュール127へアクセスし、キャッシュメモリモジュールのディレクトリ情報を更新する。具体的には、マイクロプロセッサ101は、キャッシュメモリモジュールの内容が更新されたことをディレクトリ情報に登録する(751)。更にマイクロプロセッサ101は、データ読み出しの要求コマンドを受けたインターフェース部10に対して、メモリ部21から要求データを読み出すよう、指示を送る。

#### 【0060】

指示を受けたインターフェース部10は、キャッシュヒット時の処理手順と同様に、キャッシュメモリモジュール126から要求データを読み出し、サーバ3へ転送する。以上のようにして、ストレージシステム1は、サーバ3からのデータの読み出し要求に対し、キャッシュメモリモジュールまたはハードディスク群2からデータを読み出して、サーバ3へ送信する。

#### 【0061】

次に、サーバ3からストレージシステム1にデータを書き込む場合の処理手順の例を述べる。図23は、サーバ3からストレージシステム1にデータを書き込む場合の処理手順

の例を示したフローチャートである。

まず、サーバ3は、ストレージシステム1に対してデータの書き込みコマンドを発行する。なお、本実施形態では、書き込みコマンドには、書き込まれるべきデータ（以下「更新データ」とも言う）が含まれているとして説明を行う。ただし、書き込みコマンドには更新データが含まれない場合もある。この場合は、一旦書き込みコマンドによってストレージシステム1の状態を確認した後で、サーバ3は更新データを送信する。

#### 【0062】

インターフェース部10内の外部IF100がコマンドを受信する（762）と、コマンド待ち（761）にあった外部IF100は、受信したコマンドを、転送制御部105及びスイッチ部51を介してプロセッサ部81内の転送制御部105へ転送する。転送制御部105は、受信したコマンドをプロセッサ部のメモリモジュール123へ書き込む。尚、更新データは、インターフェース部10のメモリモジュール123に一時的に保存される。

#### 【0063】

プロセッサ部81のマイクロプロセッサ101は、メモリモジュール123へコマンドが書き込まれたことを、メモリモジュール123へのポーリングや転送制御部105からの書き込みを示す割り込み等によって検出する。コマンドの書き込みを検出したマイクロプロセッサ101は、メモリモジュール123から該当するコマンドを読み出してコマンド解析を行う（763）。マイクロプロセッサ101は、コマンド解析の結果から、サーバ3が書き込みを要求する更新データを記録する記憶領域を示す情報を割り出す（764）。マイクロプロセッサ101は、更新データを書き込む記憶領域を示す情報及びプロセッサ部21内のメモリモジュール123又はメモリ部21内の制御メモリモジュール127に格納されているキャッシュメモリモジュールのディレクトリ情報に基づいて、メモリ部21内のキャッシュメモリモジュール126に、書き込み要求の対象、すなわち更新対象となるデータ（以下「更新対象データ」）が記録されているかどうかを判断する（765）。

#### 【0064】

キャッシュメモリモジュール126に更新対象データがあった場合（以下「ライトヒット」とも言う。）（766）、マイクロプロセッサ101は、インターフェース部10内の外部IF100からキャッシュメモリモジュール126へ更新データを転送するために必要な情報を、プロセッサ部81内の転送制御部105、スイッチ部51及びインターフェース部10内の転送制御部105を介して、インターフェース部10内のメモリモジュール123へ転送する。そして、マイクロプロセッサ101は、外部IF100に、サーバ3から転送された更新データをメモリ部21内のキャッシュメモリモジュール126へ書き込むように指示する（768）。

#### 【0065】

指示を受けたインターフェース部10内の外部IF100は、自インターフェース部10内のメモリモジュール123の所定の場所から更新データの転送に必要な情報を読み出す。読み出した情報をもとに、インターフェース部10内の外部IF100は、転送制御部105及びスイッチ部51を介して、メモリ部21内のメモリコントローラ125へ更新データを転送する。更新データを受信したメモリコントローラ125は、キャッシュメモリモジュール126に格納されている更新対象データを要求データで上書きする（769）。書き込み終了後、メモリコントローラ125は、指示を送信したマイクロプロセッサ101へ更新データの書き込みの終了を通知する。

#### 【0066】

キャッシュメモリモジュール126への更新データの書き込みの終了を検知したマイクロプロセッサ101は、メモリ部21内の制御メモリモジュール127へアクセスし、キャッシュメモリのディレクトリ情報を更新する（770）。具体的には、マイクロプロセッサ101は、ディレクトリ情報にキャッシュメモリモジュールの内容が更新されたことを登録する。それとともに、マイクロプロセッサ101は、サーバ3から書き込み要求を

受けた外部 I F 1 0 0 に、書き込み完了通知をサーバ 3 へ送出するように指示する (7 7 1)。その指示を受けた外部 I F 1 0 0 は、書き込み完了通知をサーバ 3 へ送出する (7 7 2)。

#### 【0067】

キャッシュメモリモジュール 1 2 6 内に更新対象データがない場合 (以下「ライトミス」とも言う。) (7 6 6)、マイクロプロセッサ 1 0 1 は、メモリ部 2 1 内の制御メモリモジュール 1 2 7 へアクセスし、キャッシュメモリモジュールのディレクトリ情報に、メモリ部 2 1 内のキャッシュメモリモジュール 1 2 6 に更新データを格納する領域を確保するための情報、具体的には空きキャッシュスロットを指定する情報を登録する (キャッシュ領域確保。) (7 6 7)。キャッシュ領域確保後、ストレージシステム 1 は、ライトヒット時と同様の制御を行う。ただし、ライトミスの場合にはキャッシュメモリモジュール 1 2 6 には更新対象データが存在しないので、メモリコントローラ 1 2 5 は、更新データを格納する場所として確保された記憶領域に更新データを格納する。

#### 【0068】

その後マイクロプロセッサ 1 0 1 は、キャッシュメモリモジュール 1 2 6 の空き容量等を判断して (7 8 1)、サーバ 3 からの書き込み要求とは非同期に、メモリ部 2 1 内のキャッシュメモリモジュール 1 2 6 に書き込まれた更新データをハードディスク群 2 に記録する処理を行う。具体的には、マイクロプロセッサ 1 0 1 は、メモリ部 2 1 内の制御メモリモジュール 1 2 7 へアクセスし、記憶領域の管理情報から、更新データを格納するハードディスク群 2 が接続されているインターフェース部 1 0 (以下「更新目的インターフェース部 1 0」とも言う) を割り出す (7 8 2)。その後、マイクロプロセッサ 1 0 1 は、キャッシュメモリモジュール 1 2 6 から更新目的インターフェース部 1 0 内の外部 I F 1 0 0 へ更新データを転送するために必要な情報を、プロセッサ部 8 1 内の転送制御部 1 0 5、スイッチ部 5 1 及びインターフェース部 1 0 内の転送制御部 1 0 5 を介して、更新目的インターフェース部 1 0 内のメモリモジュール 1 2 3 へ転送する。

#### 【0069】

その後マイクロプロセッサ 1 0 1 は、キャッシュメモリモジュール 1 2 6 から更新データを読み出して更新目的インターフェース部 1 0 の外部 I F 1 0 0 へ転送するように、更新目的インターフェース部 1 0 へ指示する。指示を受けた更新目的インターフェース部 1 0 内の外部 I F 1 0 0 は、自インターフェース部 1 0 内のメモリモジュール 1 2 3 の所定の場所から更新データの転送に必要な情報を読み出す。読み出した情報をもとに、更新目的インターフェース部 1 0 内の外部 I F 1 0 0 は、メモリ部 2 1 内のメモリコントローラ 1 2 5 に、キャッシュメモリモジュール 1 2 6 から更新データを読み出し、その更新データをメモリコントローラ 1 2 5 から更新目的インターフェース部 1 0 内の転送制御部 1 0 5 を介して外部 I F 1 0 0 へ転送するよう指示する。

#### 【0070】

指示を受けたメモリコントローラ 1 2 5 は、更新データを更新目的インターフェース部 1 0 の外部 I F 1 0 0 へ転送する (7 8 3)。更新データを受信した外部 I F 1 0 0 は、ハードディスク群 2 へ更新データを書き込む (7 8 4)。以上のようにして、サーバ 3 からのデータの書き込み要求に対し、ストレージシステム 1 は、キャッシュメモリモジュールへデータを書き込み、さらにハードディスク群 2 へデータを書き込む。

#### 【0071】

本実施例に示すストレージシステム 1 では、管理端末 6 5 をストレージシステム 1 に接続し、管理端末 6 5 から、システムの構成情報の設定、システムの立上げ/停止の制御、システム内各部の利用率、稼動状況、障害情報の収集、障害時の障害部位の閉塞/交換処理、制御プログラムの更新等を行う。ここで、システムの構成情報、利用率、稼動状況、障害情報はメモリ部 2 1 の制御メモリモジュール 1 2 7 内に格納される。ストレージシステム 1 内には内部 LAN (Local Area Network) 9 1 を設ける。各プロセッサ部 8 1 は LAN インターフェースを有し、管理端末 6 5 と各プロセッサ部 8 1 は内部 LAN 9 1 で接続されている。管理端末 6 5 は内部 LAN 経由 9 1 で、各プロセッサ部

81にアクセスし、上記の各種処理を行う。

【0072】

図14及び15は、本実施例に示す構成のストレージシステム1を筐体の実装する場合の構成例を示す図である。

【0073】

ストレージシステム1の骨格を構成する筐体は、電源ユニットシャーシ823、制御ユニットシャーシ821及びディスクユニットシャーシ822とを有する。これらのシャーシに上述した各部が装填される。制御ユニットシャーシ821の一面には、インターフェース部10、スイッチ部51、プロセッサ部81及びメモリ部21間を接続する信号線がプリントされたバックプレーン831が設けられている(図15)。バックプレーン831は各層に信号線をプリントした複数層の基板から構成される。バックプレーン831は、IFパッケージ801、SWパッケージ802、メモリパッケージ803又はプロセッサパッケージ804が接続されるコネクタ911を有する。バックプレーン831上の信号線は、各パッケージが接続されるコネクタ911内の定められた端子に繋がるようにプリントされている。また、同各パッケージへ電源を供給するための電源用の信号線がバックプレーン831上にプリントされている。

【0074】

IFパッケージ801は、各層に信号線をプリントした複数層の回路基板から構成される。IFパッケージ801は、バックプレーン831に接続するためのコネクタ912を有する。またIFパッケージ801の回路基板には、図8に示すインターフェース部10の構成における外部IF100と転送制御部105間の信号線、メモリモジュール123と転送制御部105間の信号線及び転送制御部105をスイッチ部51へ接続する信号線をコネクタ912に接続する信号線がプリントされている。更に、IFパッケージ801の回路基板上には、外部IF100の役割を果たす外部IF-LSI901、転送制御部105の役割を果たす転送制御LSI902及びメモリモジュール123を構成する複数のメモリLSI903が回路基板上の配線に従って実装されている。

【0075】

また、外部IF-LSI901、転送制御LSI902及びメモリLSI903を駆動するための電源並びにクロック用の信号線もIFパッケージ801の回路基板上にプリントされている。また、IFパッケージ801は、サーバ3またはハードディスク群2と外部IF-LSI901とを接続するためのケーブル920をIFパッケージ801に接続するためのコネクタ913を有する。コネクタ913と外部IF-LSI901間の信号線が回路基板上にプリントされている。

【0076】

SWパッケージ802、メモリパッケージ803及びプロセッサパッケージ804も、基本的にIFパッケージ801と同様の構成である。つまり、具体的には上述した各部の役割を果たすLSIが回路基板の上に実装されており、その間を接続する信号線が回路基板にプリントされている。ただし、その他のパッケージは、IFパッケージ801が備えるコネクタ913及びそれに接続するための信号線は備えていない。

【0077】

制御ユニットシャーシ821の上には、ハードディスクドライブを実装したハードディスクユニット811を装填するためのディスクユニットシャーシ822を設ける。ディスクユニットシャーシ822は、ハードディスクユニット811とディスクユニットシャーシとを接続するためのバックプレーン832を有する。ディスクユニット811とバックプレーン832は、両者を接続するためのコネクタを有する。バックプレーン831と同様に、バックプレーン832は各層に信号線をプリントした複数層の基板から構成されている。更にバックプレーン832は、IFパッケージ801に接続されるケーブル920が接続されるコネクタを有する。そのコネクタとディスクユニット811を接続するコネクタとの間の信号線及び電源供給用の信号線がバックプレーン832にプリントされている。



**【0078】**

また、ケーブル 920 を接続する専用のパッケージを設け、そのパッケージをバックプレーン 832 に設けたコネクタに接続しても良い。

**【0079】**

また、制御ユニットシャーシ 821 の下には、ストレージシステム 1 全体に電力を供給する電源ユニットやバッテリーユニットを収めた電源ユニットシャーシ 823 を設ける。

そして、これらのシャーシを 19 インチラック（図示していない）に収める。尚、シャーシの配置関係は図示した例に限られず、例えば電源ユニットシャーシが筐体の一番上に装填されても良い。

**【0080】**

なお、ストレージシステム 1 はハードディスク群 2 を有さない構成もありうる。この場合、ストレージシステム 1 とは別の場所に存在するハードディスク群 2 や他のストレージシステム 1 とストレージシステム 1 とが I F パッケージ 801 に設けられた接続ケーブル 920 を介して接続されることになる。またこの場合、ハードディスク群 2 はディスクユニットシャーシ 822 に収め、ディスクユニットシャーシ 822 をディスクユニットシャーシ専用の 19 インチラックに収める。更にストレージシステム 1 はハードディスク群 2 を有し、さらに他のストレージシステム 1 と接続する場合もある。この場合も、ストレージシステム 1 と他のストレージシステム 1 は I F パッケージ 801 に設けられた接続ケーブル 920 を介して相互に接続されることになる。

**【0081】**

また上記では、インターフェース部 10、プロセッサ部 81、メモリ部 21 及びスイッチ部をそれぞれ別個のパッケージに実装する場合について説明したが、例えば、スイッチ部 51、プロセッサ部 81 及びメモリ部 21 をまとめて 1 枚のパッケージに実装することも可能である。また、インターフェース部 10、スイッチ部 51、プロセッサ部 81 及びメモリ部 21 全てをまとめて 1 枚のパッケージに実装することも可能である。そうした場合、パッケージのサイズが変わり、それに合わせて図 18 に示す制御ユニットシャーシ 821 の幅、高さを変える必要がある。また図 14 では、パッケージを床面と垂直になる形態で制御ユニットシャーシ 821 へ実装しているが、パッケージが床面と水平になる形態で制御ユニットシャーシ 821 へ実装することも可能である。上記のインターフェース部 10、プロセッサ部 81、メモリ部 21 及びスイッチ部 51 の内、どの組合せを 1 つのパッケージに実装するかは任意であり、上記の実装の組合せは一例である。

**【0082】**

制御ユニットシャーシ 821 に搭載可能なパッケージの枚数は、制御ユニットシャーシ 821 の幅と各パッケージの厚さから物理的に決まる。一方、図 2 に示す構成からわかるように、ストレージシステム 1 は、スイッチ部 51 を介して、インターフェース部 10、プロセッサ部 81 及びメモリ部 21 を互いに接続する構成であるため、要求されるシステム規模、サーバ接続数、ハードディスク接続数、性能に合わせて、各部の数を自由に設定可能である。したがって、図 14 に示す I F パッケージ 801、メモリパッケージ 803 及びプロセッサパッケージ 804 に設けるバックプレーン 831 とのコネクタを共通化し、さらに、搭載する S W パッケージ 802 の枚数と S W パッケージ 802 を接続するバックプレーン 831 上のコネクタとを予め定めておくことにより、制御ユニットシャーシ 821 に搭載可能なパッケージの枚数から搭載する S W パッケージの枚数を差し引いた数を上限として、I F パッケージ 801、メモリパッケージ 803 及びプロセッサパッケージ 804 の枚数を自由に選んで搭載することが可能となる。こうすることにより、ユーザが要求するシステム規模、サーバ接続数、ハードディスク接続数、性能に合わせて、フレキシブルにストレージシステム 1 を構成することが可能となる。

**【0083】**

本実施例では、図 20 に示す従来技術のチャンネル I F 部 11 及びディスク I F 部 16 からマイクロプロセッサ 103 を分離し、プロセッサ部 81 として独立させることに特徴がある。こうすることにより、サーバ 3 あるいはハードディスク群 2 との接続インターフェ

ース数の増減とは独立にマイクロプロセッサ数の増減が可能となり、サーバ3やハードディスク群2の接続数やシステムの性能といったユーザの要求に柔軟に応えることが可能なフレキシブルな構成のストレージシステムの提供を可能にできる。

#### 【0084】

また本実施例では、データのリードまたはライト時に、チャンネルIF部11内のマイクロプロセッサ103で行っていた処理と、ディスクIF部16内のマイクロプロセッサ103で行っていた処理を、図1に示すプロセッサ部81内の1つのマイクロプロセッサ101でまとめて一貫して処理する。こうすることにより、従来技術において必要であったチャンネルIF部とディスクIF部のそれぞれのマイクロプロセッサ103間の処理の引継ぎのオーバーヘッドを削減することが可能となる。

#### 【0085】

また、プロセッサ部81の2つのマイクロプロセッサ101、あるいは異なるプロセッサ部81各々から1つずつ選択される2つのマイクロプロセッサ101により、そのうちの一方のマイクロプロセッサ101がサーバ3とのインターフェース部10側の処理、もう一方がハードディスク群2とのインターフェース部10側の処理を行っても良い。

#### 【0086】

また、サーバ3とのインターフェース側の処理の負荷がハードディスク群2とのインターフェース側の処理の負荷より大きい場合、前者の処理により多くのマイクロプロセッサ101の処理量（例えばプロセッサ数、一つのプロセッサの占有率等）を割り当てることができる。負荷の大きさが逆の場合は、後者の処理により多くのマイクロプロセッサ101の処理量を割り当てることができる。したがって、ストレージシステム内の各処理の負荷の大小により、柔軟にマイクロプロセッサの処理量（リソース）を割り当てることが可能となる。

#### 【実施例2】

#### 【0087】

図5は、第二の実施例の構成例を示す図である。

ストレージシステム1は、複数のクラスタ70-1~70-nを相互結合網31で相互に接続した構成を有する。一つのクラスタ70は、サーバ3またはハードディスク群2が繋がるインターフェース部10、メモリ部21、プロセッサ部81及び相互結合網31の一部を、あるまとまった数有する。一つのクラスタ70が有する各部の数は任意である。各クラスタ70のインターフェース部10、メモリ部21及びプロセッサ部81は相互結合網31に接続される。したがって、各クラスタ70の各部は、他のクラスタ70の各部と相互結合網31を介してパケットのやり取りを行うことができる。尚、各クラスタ70は、ハードディスク群2を有していても良い。したがって、一つのストレージシステム1の中に、ハードディスク群2を有するクラスタ70とハードディスク群2を有さないクラスタ70とが混在する場合もある。又、全てのクラスタ70がハードディスク群2を有する場合もある。

#### 【0088】

図6は、相互結合網31の具体的な構成例を示す図である。

相互結合網31は、4つのスイッチ部51及びこれらを接続する通信バスを有する。これらのスイッチ51は個々のクラスタ70の内部に設置される。ストレージシステム1は2つのクラスタ70を有する。1つのクラスタ70は、4つのインターフェース部10、2つのプロセッサ部81及びメモリ部21を有する。また上述したように、一つのクラスタ70には、相互結合網31であるスイッチ51のうち二つが含まれる。

#### 【0089】

インターフェース部10、プロセッサ部81及びメモリ部21は、各部を含むクラスタ70内の2つのスイッチ部51と1本ずつの通信バスで接続される。これにより、インターフェース部10、プロセッサ部81及びメモリ部21の間で2つの通信経路を確保し、信頼性をあげることが可能となる。

#### 【0090】

また、クラスタ70-1とクラスタ70-2とを接続するため、1つのクラスタ70内の1つのスイッチ部51は他のクラスタ70内の2つのスイッチ部51と各々1本ずつの通信パスで接続されている。これにより、1つのスイッチ部51の故障あるいはスイッチ部51間の通信パスの故障の際にもクラスタを跨ったアクセスが可能となり、信頼性を向上することができる。

#### 【0091】

図7はストレージシステム1内におけるクラスタ間接続の異なる形態の例を示す図である。図7に示すように、クラスタ間接続専用のスイッチ部55で各クラスタ70間を接続する。この場合、クラスタ70-1～3の各スイッチ部51は2つのスイッチ部55に各々1本ずつの通信パスで接続される。これにより、1つのスイッチ部55の故障、あるいはスイッチ部51-スイッチ部55間の通信パスの故障の際にもクラスタを跨ったアクセスが可能となり、信頼性を向上することができる。

#### 【0092】

またこの場合、図6の構成に比べてクラスタの接続数を増やすことが可能となる。すなわち、スイッチ部51に接続可能な通信パスの数は物理的に限界がある。しかし専用のスイッチ部55をクラスタ間接続に使用することにより、図6の構成に比べてクラスタの接続数を増やすことが可能となる。

#### 【0093】

本実施例の構成においても、図20に示す従来技術において、チャンネルIF部11及びディスクIF部16からマイクロプロセッサ103を分離し、プロセッサ部81内に独立させるということに特徴がある。こうすることにより、サーバ3あるいはハードディスク群2との接続インターフェース数の増減とは独立にマイクロプロセッサ数の増減が可能となり、サーバ3やハードディスク群2の接続数やシステムの性能といったユーザの要求に柔軟に応えることが可能なフレキシブルな構成のストレージシステムを提供することができる。

#### 【0094】

また本実施例でも、第一の実施例と同様のデータのリード及びライト処理が行われる。したがって、本実施例でも、データのリードまたはライト時に、チャンネルIF部11内のマイクロプロセッサ103で行っていた処理と、ディスクIF部16内のマイクロプロセッサ103で行っていた処理を、図1に示すプロセッサ部81内の1つのマイクロプロセッサ101でまとめて一貫して処理する。こうすることにより、従来技術において必要であったチャンネルIF部とディスクIF部それぞれのマイクロプロセッサ103間の処理の引継ぎのオーバーヘッドを削減することが可能となる。

#### 【0095】

尚、本実施例でデータのリード又はライトを実行する場合、一つのクラスタ70に接続されたサーバ3から他のクラスタ70が有するハードディスク群2（あるいは他のクラスタ70に接続されたストレージシステム）へのデータのライト又はリードが行われる場合がある。この場合でも、第一の実施例で説明したリード及びライト処理が行われる。この場合、個々のクラスタ70が有するメモリ部21のメモリ空間をストレージシステム1全体で一つの論理的なメモリ空間とすることで、一つのクラスタのプロセッサ部81等は、他のクラスタ70のメモリ部21等へアクセスするための情報を得ることが出来る。又、一つのクラスタのプロセッサ部81は、他のクラスタが有するインターフェース部10に対してデータの転送を指示することが出来る。

#### 【0096】

又、ストレージシステム1は、各クラスタに接続されるハードディスク群2から構成されるボリュームを、全てのプロセッサ部で共有されるように、一つのメモリ空間で管理する。

#### 【0097】

また本実施例でも第一の実施例と同様に、管理端末65をストレージシステム1に接続し、管理端末65から、システムの構成情報の設定、システムの立上げ/停止の制御、シ

システム内各部の利用率、稼動状況、障害情報の収集、障害時の障害部位の閉塞／交換処理、制御プログラムの更新等を行う。ここで、システムの構成情報、利用率、稼動状況、障害情報はメモリ部21の制御メモリモジュール127内に格納される。本実施例の場合、ストレージシステム1を複数のクラスタ70から構成するため、補助プロセッサを備えたボード（補助プロセッサ部85）をクラスタ70毎に設ける。補助プロセッサ部85は、管理端末65の指示を各プロセッサ部81に伝えたり、各プロセッサ部81からの情報を纏めて管理端末65へ送る役目を果たす。管理端末65と補助プロセッサ部85を内部LAN92で接続する。そして、クラスタ70内では内部LAN91を設け、各プロセッサ部81はLANインターフェースを有し、補助プロセッサ部85と各プロセッサ部81とは内部LAN91で接続される。管理端末65は補助プロセッサ部85経由で各プロセッサ部81にアクセスし、上記の各種処理を行う。尚、補助プロセッサ無しで直接プロセッサ部81と管理端末65がLAN等を介して接続されていても良い。

#### 【0098】

図17は、ストレージシステム1の本実施例における更なる変形例である。図17に示すように、サーバ3またはハードディスク群2を繋ぐインターフェース部10に他のストレージシステム4が接続される。この場合、ストレージシステム1は、他のストレージシステム4が繋がるインターフェース部10が属するクラスタ70内の制御メモリモジュール126及びキャッシュメモリモジュール127に、他のストレージシステム4が提供する記憶領域（以下「ボリューム」とも言う）の情報及び他のストレージシステム4に格納される（又は読み出される）データを格納する。

#### 【0099】

他のストレージシステム4が接続されるクラスタ70内のマイクロプロセッサ101は、制御メモリモジュール127に格納された情報に基づいて、他のストレージシステム4が提供するボリュームを管理する。例えばマイクロプロセッサ101は、他のストレージシステム4が提供するボリュームをサーバ3にストレージシステム1が提供するボリュームとして割り当てる。このことにより、サーバ3が、ストレージシステム1を介して他のストレージシステム4のボリュームへアクセスすることが可能となる。

この場合、ストレージシステム1は、自己が有するハードディスク群2から構成されるボリュームと他のストレージシステム4が提供するボリュームを一括して管理する。

#### 【0100】

また、図17において、ストレージシステム1が、どのインターフェース部10にどのサーバ3が接続されているかを示すテーブルをメモリ部21内の制御メモリモジュール127に格納する。そして同じクラスタ70内のマイクロプロセッサ101がそのテーブルを管理する。具体的には、サーバ3とホストIF100との接続関係が追加変更等された場合、マイクロプロセッサ101は、上述したテーブルの内容を変更（更新、追加又は削除）する。これにより、ストレージシステム1に繋がる複数のサーバ3間でのストレージシステム1を介した通信およびデータ転送が可能となる。このことは、第一の実施例でも同様に実現可能である。

#### 【0101】

更に図17において、インターフェース部10に繋がるサーバ3がストレージシステム4との間でデータ転送を行う際、ストレージシステム1は、サーバ3が接続されるインターフェース部10とストレージシステム4が接続されるインターフェース部10との間で相互結合網31を介してデータ転送を行う。この時、ストレージシステム1は、転送されるデータをメモリ部21内のキャッシュメモリモジュール126にキャッシングしても良い。これにより、サーバ3およびストレージシステム4間のデータ転送性能が向上する。

#### 【0102】

また本実施例において、図18に示すように、スイッチ65を介して、ストレージシステム1とサーバ3および他のストレージシステム4との間を接続する構成も考えられる。この場合、サーバ3は、インターフェース部10内の外部IF100及びスイッチ65を介してサーバ3および他のストレージシステム4へアクセスする。こうすることにより、

ストレージシステム 1 に接続されたサーバ 3 から、スイッチ 65 および複数のスイッチ 65 からなるネットワークに接続されたサーバ 3 や他のストレージシステム 4 へのアクセスが可能となる。

#### 【0103】

図 19 は、図 6 に示す構成のストレージシステム 1 を筐体の実装する場合の構成例を示す図である。

実装の構成は基本的に図 14 の実装構成と同様である。すなわち、インターフェース部 10、プロセッサ部 81、メモリ部 21 及びスイッチ部 51 をパッケージに実装し、制御ユニットシャーシ 821 内のバックプレーン 831 に接続する。

#### 【0104】

図 6 の構成では、インターフェース部 10、プロセッサ部 81、メモリ部 21 及びスイッチ部 51 がクラスタ 70 としてグループ化される。そこで、クラスタ 70 毎に 1 つの制御ユニットシャーシ 821 を用意する。1 つのクラスタ 70 内の各部は、1 つの制御ユニットシャーシ 821 に実装される。すなわち、異なるクラスタ 70 のパッケージを異なる制御ユニットシャーシ 821 に実装する。更に、クラスタ 70 間の接続のため、図 19 に示すように異なる制御ユニットシャーシに装填された SW パッケージ 802 間をケーブル 921 で接続する。この場合、図 19 に示す IF パッケージ 801 と同様に、SW パッケージ 802 にはケーブル 921 接続用のコネクタが実装される。

#### 【0105】

なお、1 つの制御ユニットシャーシ 821 に実装されるクラスタの数は、1 つで無くても良い。例えば、1 つの制御ユニットシャーシ 821 に実装されるクラスタ数は 2 つであっても良い。

#### 【実施例 3】

#### 【0106】

実施例 1 及び 2 の構成のストレージシステム 1 においては、インターフェース部 10 で受信するコマンドの解析をプロセッサ部 81 で行う。しかしながら、サーバ 3 とストレージシステム 1 との間でやり取りされるコマンドが従うプロトコルは多種多様あり、一般的なプロセッサで全てのプロトコルの解析処理を行うことは現実的ではない。ここで、プロトコルとは、例えばファイル名を用いたファイル I/O (Input/Output) プロトコル、iSCSI (internet Small Computer System Interface) プロトコル、サーバとして大型計算機 (メインフレーム) を使用する際のプロトコル (チャンネルコマンドワード: CCW) 等がある。

#### 【0107】

そこで、本実施例では、これらのプロトコルを高速に処理する専用のプロセッサを実施例 1 及び 2 の全て又は一部のインターフェース部 10 に追加する。図 13 は、転送制御部 105 にマイクロプロセッサ 102 を接続したインターフェース部 10 (以下このインターフェース部 10 を「アプリケーション制御部 19」と称する) の例を示す図である。

#### 【0108】

本実施例のストレージシステム 1 は、実施例 1 及び 2 のストレージシステム 1 が有する全ての又は一部のインターフェース部 10 の代わりにアプリケーション制御部 19 を有する。アプリケーション制御部 19 は、相互結合網 31 と接続される。ここで、アプリケーション制御部 19 が有する外部 IF 100 は、アプリケーション制御部 19 のマイクロプロセッサ 102 が処理するプロトコルに従うコマンドを専用受信する外部 IF とする。ただし、1 つの外部 IF 100 で異なるプロトコルに従う複数のコマンドを受信する構成としても良い。

#### 【0109】

マイクロプロセッサ 102 は、プロトコル変換処理を外部 IF 100 と連携して行う。具体的には、マイクロプロセッサ 102 は、サーバ 3 からアプリケーション制御部 19 がアクセス要求を受けた場合、外部 IF が受信したコマンドのプロトコルを内部のデータ転送用のプロトコルに変換する処理を行う。

**【0110】**

また、専用のアプリケーション制御部19を用意する代わりに、インターフェース部10はそのまま使用し、プロセッサ部81内のマイクロプロセッサ101のうちの一つをプロトコル処理専用とする構成も考えられる。

**【0111】**

本実施例におけるデータのリード及びライト処理は、第一の実施例と同様に行われる。ただし、第一の実施例ではコマンドを受信したインターフェース部10は、コマンドを解析せずにプロセッサ部81へ転送していたが、本実施例では、アプリケーション処理部19においてコマンドの解析処理を行う。そして、その解析結果（コマンドの内容、データのあて先等）をアプリケーション制御部19がプロセッサ部81へ転送する。プロセッサ部81は、解析された情報に基づいて、ストレージシステム1内でのデータ転送の制御を行う。

**【0112】**

尚、本発明の他の実施態様としては、以下のような構成も考えられる。具体的には、計算機またはディスク装置とのインターフェースを有する複数のインターフェース部、計算機またはディスク装置との間でリード／ライトされるデータを格納するキャッシュメモリとシステムの制御情報を格納する制御メモリを有する複数のメモリ部、計算機とディスク装置との間でのデータのリード／ライトを制御するマイクロプロセッサを有する複数のプロセッサ部を有し、複数のインターフェース部、複数のメモリ部及び複数のプロセッサ部は少なくとも1つのスイッチ部からなる相互結合網で互いに接続されており、相互結合網を介して複数のインターフェース部、複数のメモリ部及び複数のプロセッサ部の間でデータあるいは制御情報の送受信を行うストレージシステムである。

**【0113】**

そして、本構成において、インターフェース部、メモリ部及びプロセッサ部は、データあるいは制御情報の送受信を制御する転送制御部を有する。また、本構成において、インターフェース部は第1の回路基板上に、メモリ部は第2の回路基板上に、プロセッサ部は第3の回路基板上に、少なくとも1つのスイッチ部は第4の回路基板上に実装される。更に、本構成では、第1～4の回路基板間を接続する信号線がプリントされ、プリントされた信号線に前記第1～4の回路基板を接続するための第1のコネクタを備える少なくとも1つのバックプレーンを有する。更に本構成では、第1～4の回路基板は前記バックプレーンの第1のコネクタに接続するための第2のコネクタを備えている。

**【0114】**

尚、上述した実施態様において、バックプレーンに接続できる回路基板の総数を $n$ とし、第4の回路基板の数と接続場所を予め定め、1～4の回路基板の総数が $n$ を超えない範囲でバックプレーンに接続する前記第1、第2、及び第3の回路基板それぞれの数を自由に選択できるようにしても良い。

**【0115】**

又、本発明の別の実施態様としては、以下のような構成も考えられる。具体的には、計算機またはディスク装置とのインターフェースを有する複数のインターフェース部、計算機またはディスク装置との間でリード／ライトされるデータを格納するキャッシュメモリとシステムの制御情報を格納する制御メモリを有する複数のメモリ部及び計算機とディスク装置との間でのデータのリード／ライトを制御するマイクロプロセッサを有する複数のプロセッサ部を有するクラスタを複数有するストレージシステムである。

**【0116】**

本構成において、各クラスタが有する複数のインターフェース部、複数のメモリ部及び複数のプロセッサ部の間は複数のスイッチ部からなる相互結合網で複数のクラスタを跨って互いに接続される。これにより、相互結合網を介して各クラスタ間で複数のインターフェース部、複数のメモリ部及び複数のプロセッサ部の間でデータあるいは制御情報の送受信を行う。又、本構成において、インターフェース部、メモリ部及びプロセッサ部は、それぞれスイッチに接続され、データあるいは制御情報の送受信を制御する転送制御部を有

する。

【0117】

更に本構成において、インターフェース部は第1の回路基板上に、メモリ部は第2の回路基板上に、プロセッサ部は第3の回路基板上に、少なくとも1つのスイッチ部は第4の回路基板上に実装される。そして、本構成は、第1～4の回路基板間を接続する信号線がプリントされ、プリントされた信号線に第1～4の回路基板を接続するための第1のコネクタを備える複数のバックプレーンを有し、第1～4の回路基板は前記バックプレーンの第1のコネクタに接続するための第2のコネクタを有する。本構成において、クラスタは、第1～4の回路基板を接続したバックプレーンから構成される。尚、クラスタ数とバックプレーンの数を等しくする構成としても良い。

【0118】

更に本構成において、第4の回路基板はケーブルを接続するための第3のコネクタを備え、第3のコネクタとスイッチ部を接続する信号線を第4の基板上に配線する。このようにすると、クラスタ間は、第3のコネクタ間をケーブルによって接続することで接続される。

【0119】

更に、本発明の別の実施態様としては、以下のような構成も考えられる。具体的には、計算機またはディスク装置とのインターフェースを有するインターフェース部、計算機またはディスク装置との間でリード／ライトされるデータを格納するキャッシュメモリとシステムの制御情報を格納する制御メモリを有するメモリ部及び計算機とディスク装置との間でのデータのリード／ライトを制御するマイクロプロセッサを有するプロセッサ部を有し、インターフェース部、メモリ部及びプロセッサ部の間は少なくとも1つのスイッチ部からなる相互結合網で互いに接続されているストレージシステムである。本構成において、相互結合網を介してインターフェース部、メモリ部及びプロセッサ部の間でデータあるいは制御情報の送受信を行う。

【0120】

本構成において、インターフェース部は第1の回路基板上に、メモリ部、プロセッサ部及びスイッチ部は第5の回路基板上に実装される。そして本構成において、第1と第5の回路基板間を接続する信号線がプリントされ、プリントされた信号線に第1と第5の回路基板を接続するための第4のコネクタを備える少なくとも1つのバックプレーンを有し、第1と第5の回路基板はバックプレーンの第4のコネクタに接続するための第5のコネクタを備えている。

【0121】

又、本発明の別の実施態様としては、以下のような構成も考えられる。具体的には、計算機またはディスク装置とのインターフェースを有するインターフェース部、計算機またはディスク装置との間でリード／ライトされるデータを格納するキャッシュメモリとシステムの制御情報を格納する制御メモリを有するメモリ部及び計算機とディスク装置との間でのデータのリード／ライトを制御するマイクロプロセッサを有するプロセッサ部を有し、インターフェース部、メモリ部及びプロセッサ部の間は少なくとも1つのスイッチ部からなる相互結合網で互いに接続されるストレージシステムである。本構成において、インターフェース部、メモリ部、プロセッサ部及びスイッチ部は、第6の回路基板上に実装される。

【図面の簡単な説明】

【0122】

【図1】 ストレージシステム1の構成例を示す図である。

【図2】 ストレージシステム1の相互結合網の詳細構成例を示す図である。

【図3】 ストレージシステム1の他の構成例を示す図である。

【図4】 図3に示す相互結合網の詳細構成例を示す図である。

【図5】 ストレージシステムの構成例を示す図である。

【図6】 ストレージシステムの相互結合網の詳細構成例を示す図である。

- 【図 7】 ストレージシステムの相互結合網の他の詳細構成例を示す図である。
- 【図 8】 インターフェース部の構成例を示す図である。
- 【図 9】 プロセッサ部の構成例を示す図である。
- 【図 1 0】 メモリ部の構成例を示す図である。
- 【図 1 1】 スイッチ部の構成例を示す図である。
- 【図 1 2】 パケットフォーマットの一例を示す図である。
- 【図 1 3】 アプリケーション制御部の構成例を示す図である。
- 【図 1 4】 ストレージシステムの筐体への実装例を示す図である。
- 【図 1 5】 パッケージとバックプレーンの構成例を示す図である。
- 【図 1 6】 相互結合網の他の詳細構成例を示す図である。
- 【図 1 7】 インターフェース部と外部装置との接続構成例を示す図である。
- 【図 1 8】 インターフェース部と外部装置との他の接続構成例を示す図である。
- 【図 1 9】 ストレージシステムの筐体への他の実装例を示す図である。
- 【図 2 0】 従来のストレージシステムの構成例を示す図である。
- 【図 2 1】 従来のストレージシステムの他の構成例を示す図である。
- 【図 2 2】 ストレージシステム 1 のリード動作フローを示す図である。
- 【図 2 3】 ストレージシステム 1 のライト動作フローを示す図である。

【符号の説明】

【0 1 2 3】

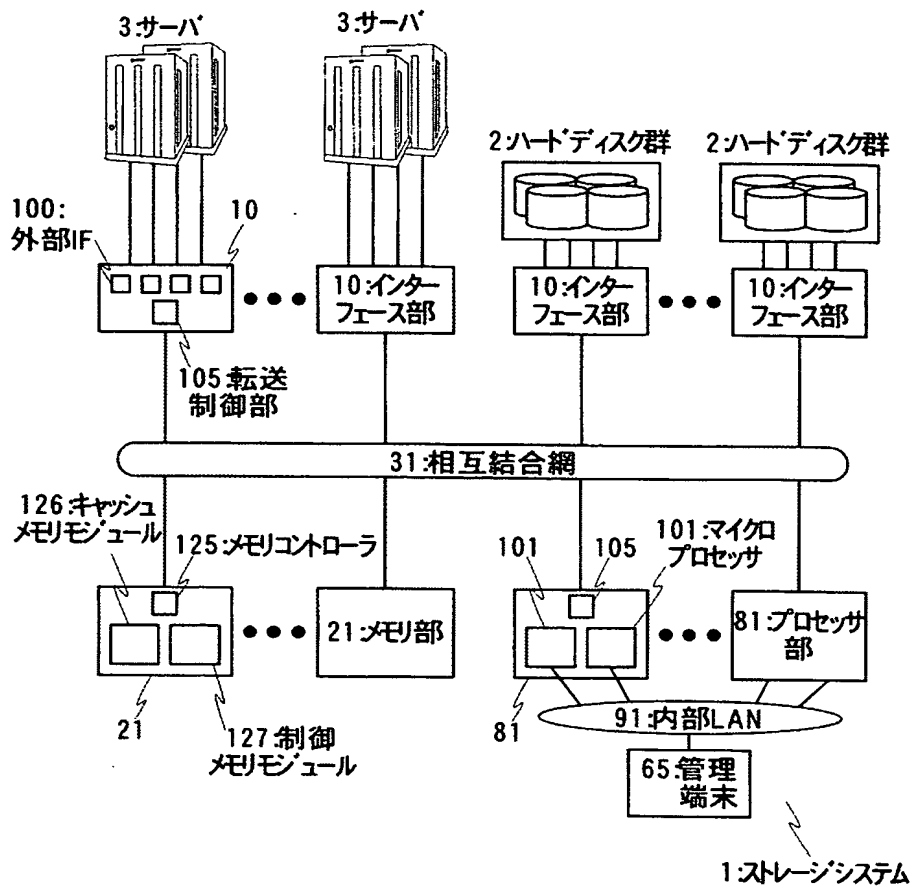
1…ストレージシステム、2…ハードディスク群、3…サーバ、1 0…インターフェース部、2 1…メモリ部、3 1…相互結合網、8 1…プロセッサ部、1 0 0…外部 I F、1 0 1…マイクロプロセッサ、1 0 5…転送制御部、1 2 5…メモリコントローラ、1 2 6…キャッシュメモリモジュール、1 2 7…制御メモリモジュール。



【書類名】 図面

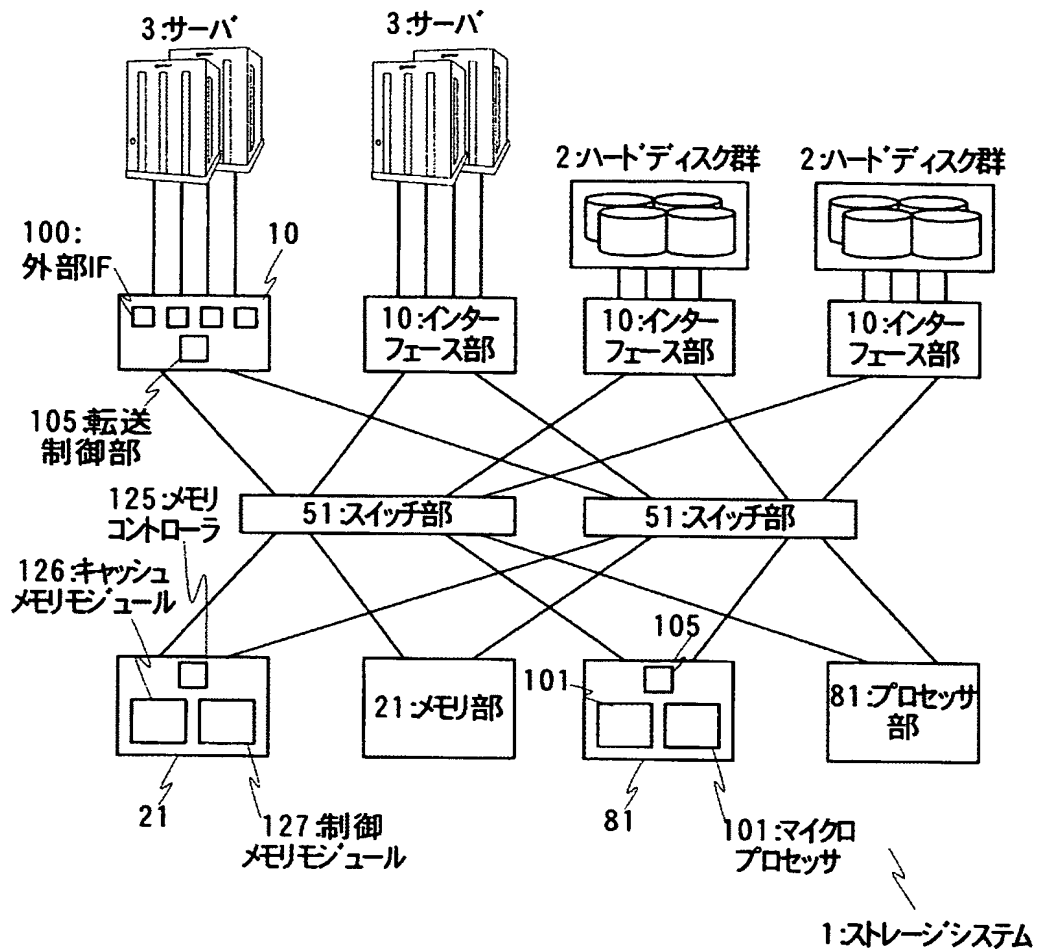
【図 1】

図1



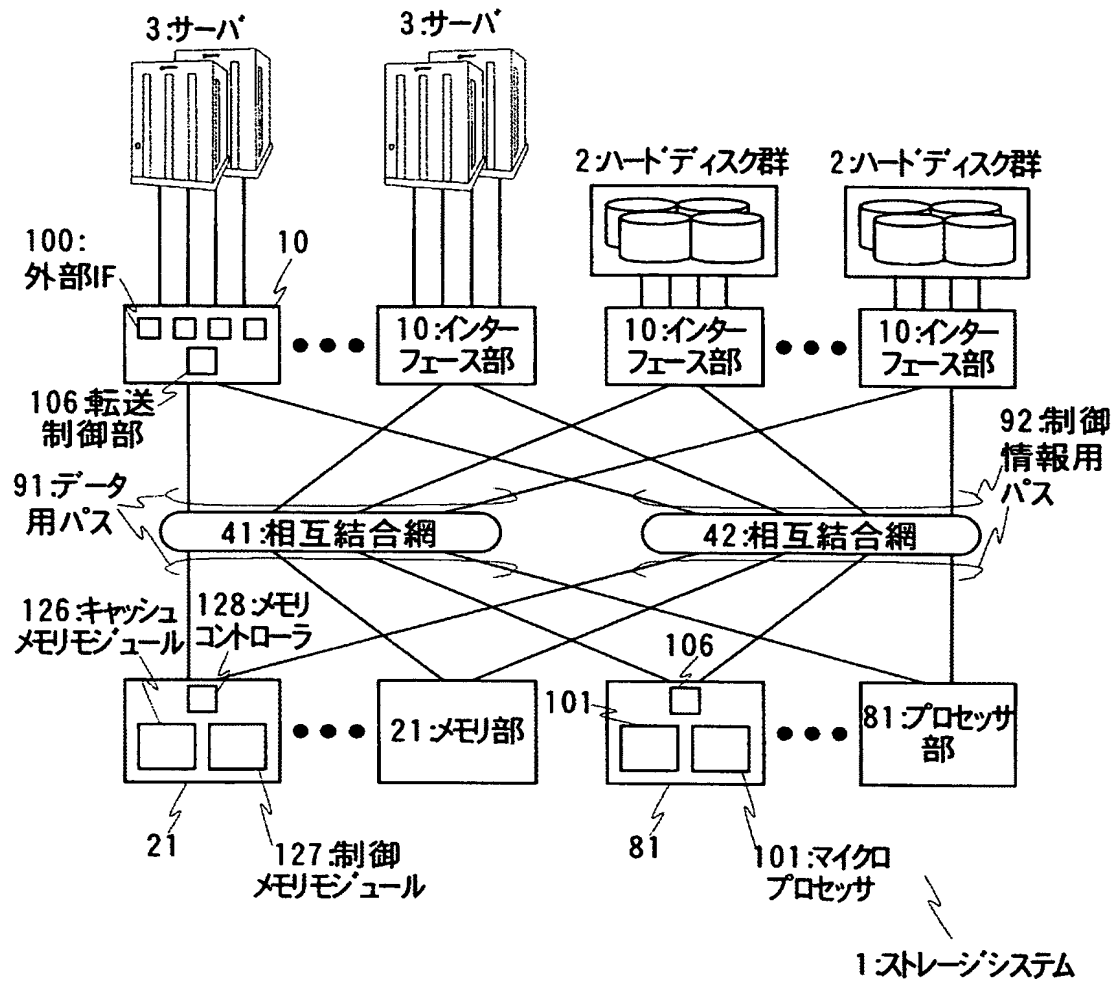
【図 2】

図 2



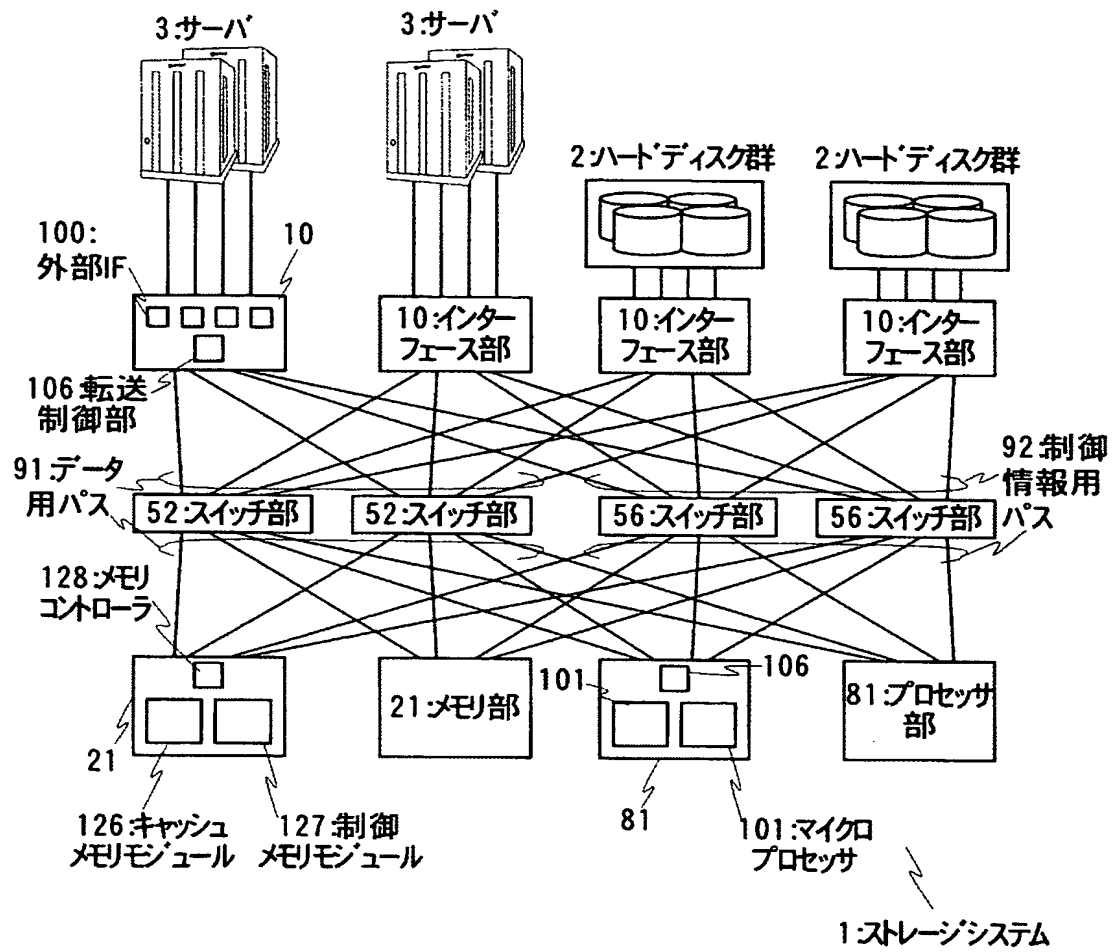
【図 3】

図 3



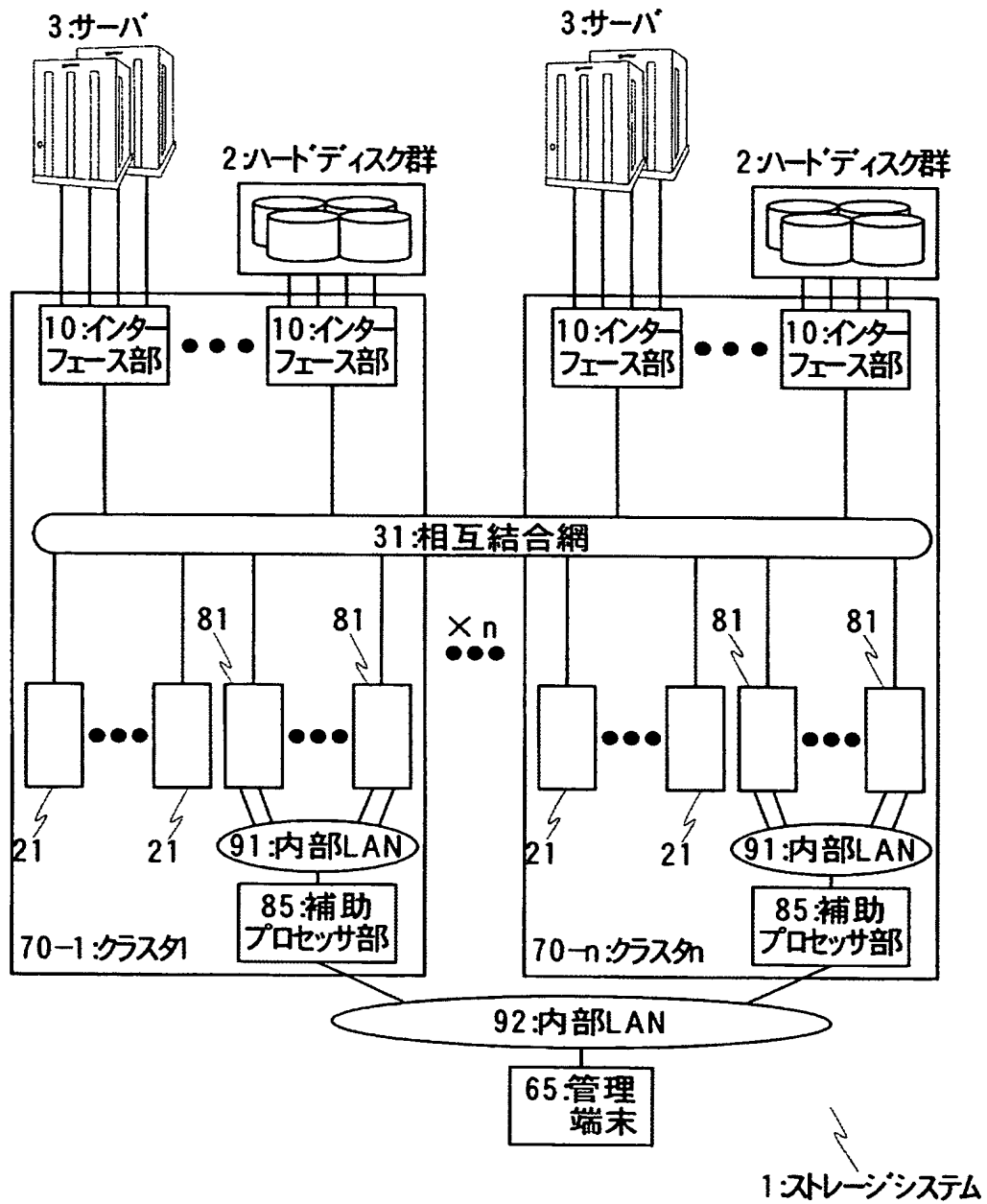
【図 4】

図 4



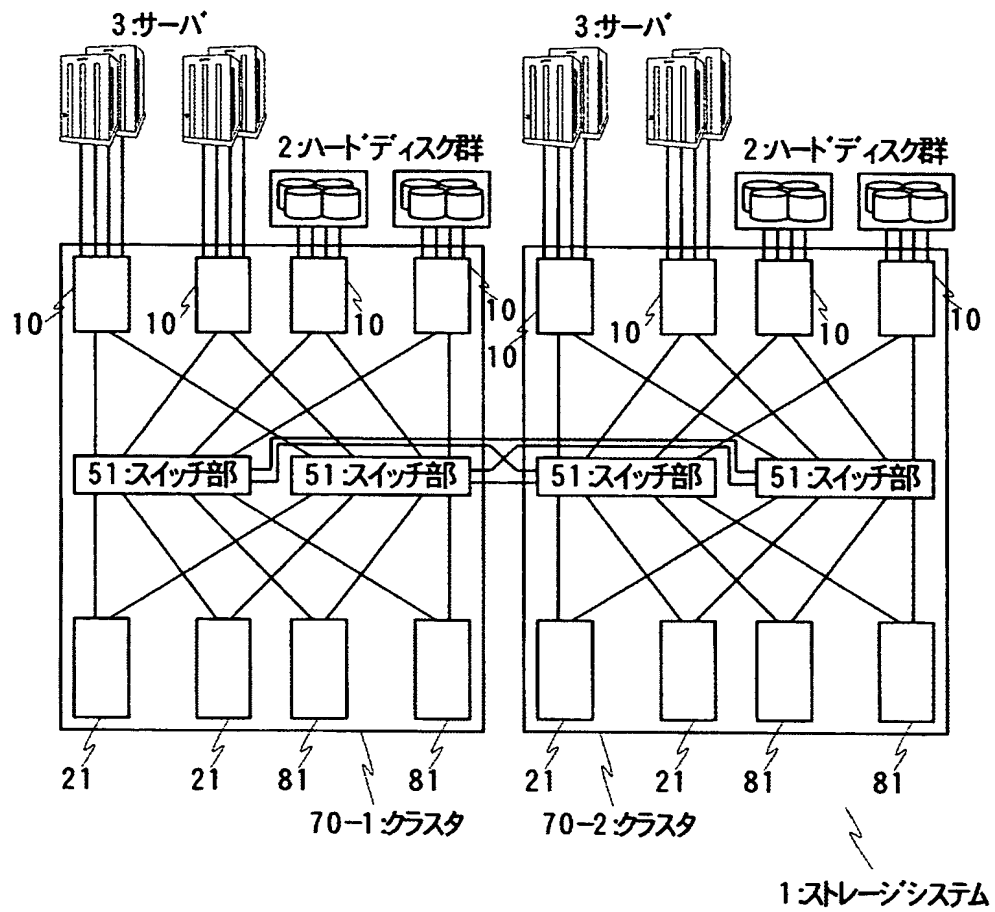
【図 5】

図5



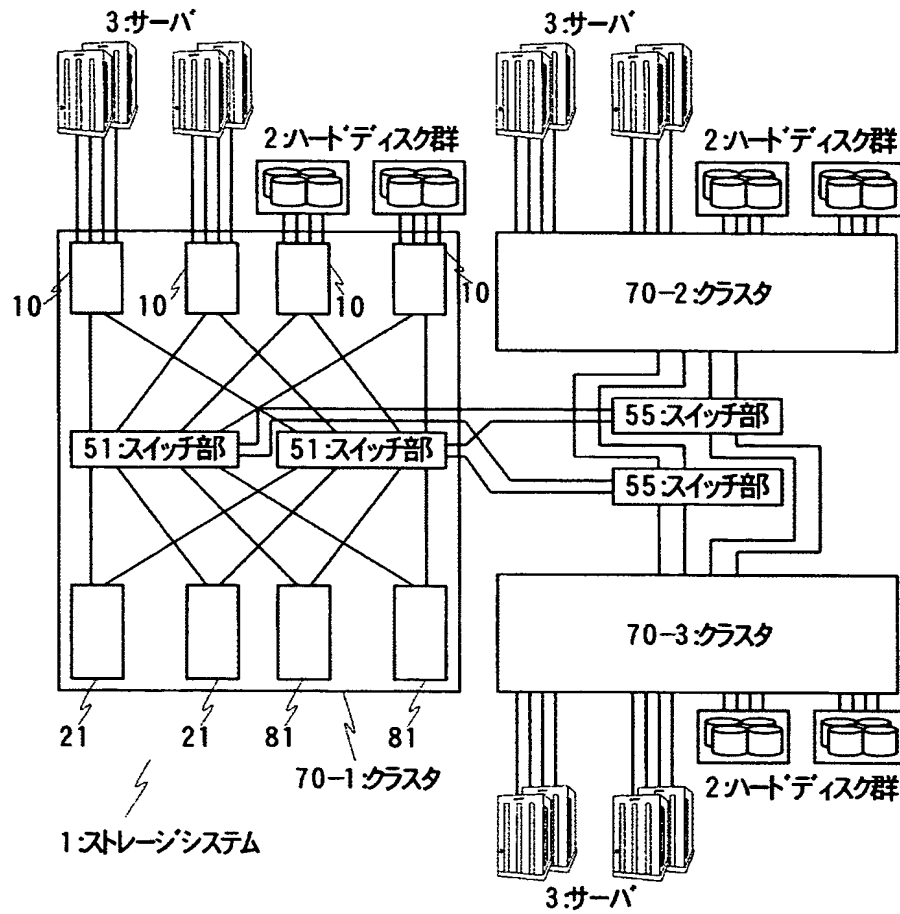
【図 6】

図6



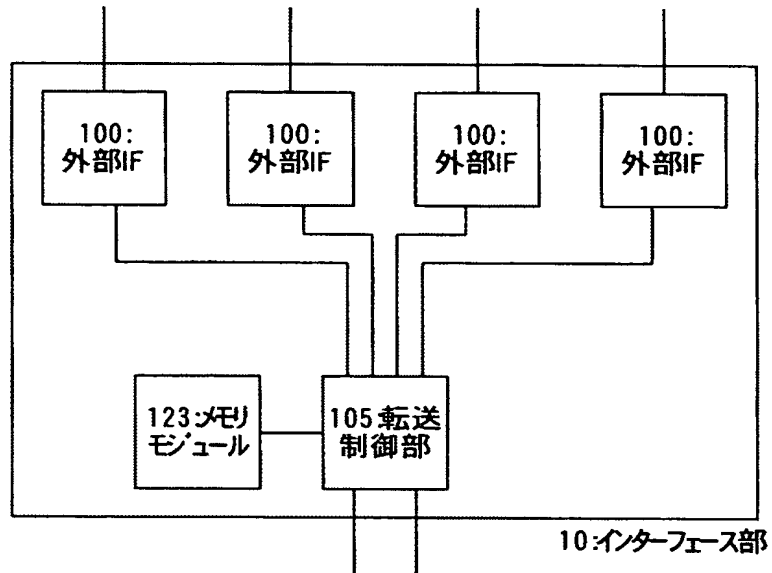
【図 7】

図 7



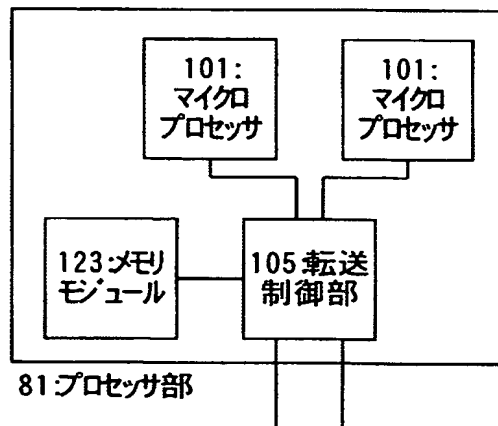
【図 8】

図8



【図 9】

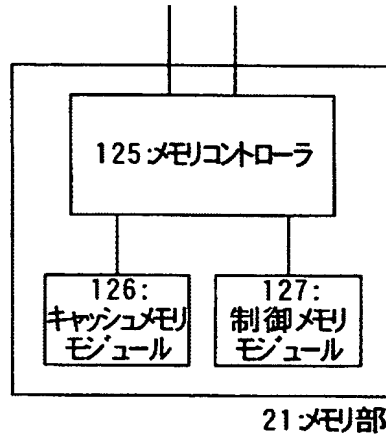
図9





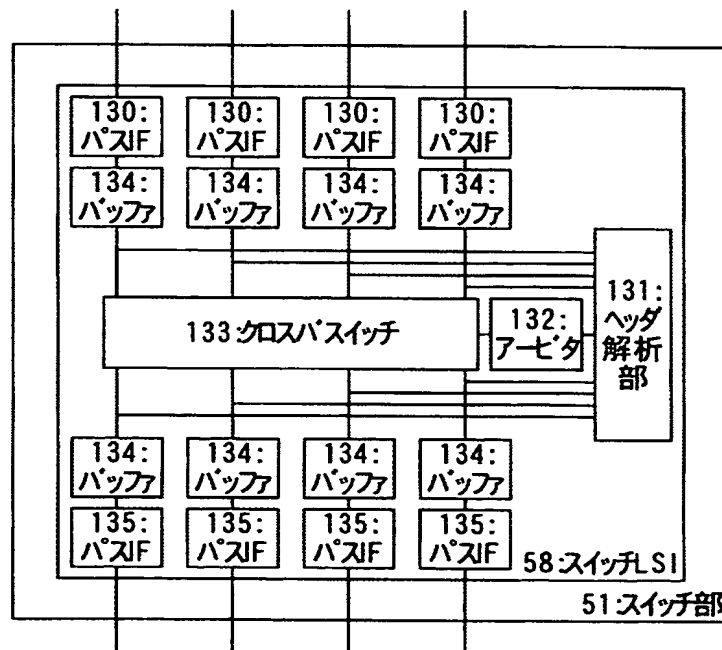
【図 10】

図10



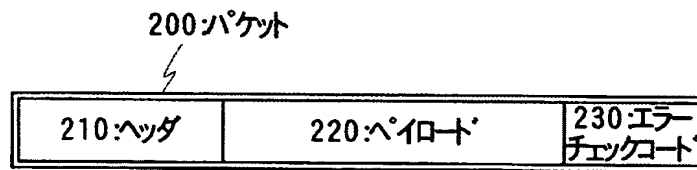
【図 11】

図11



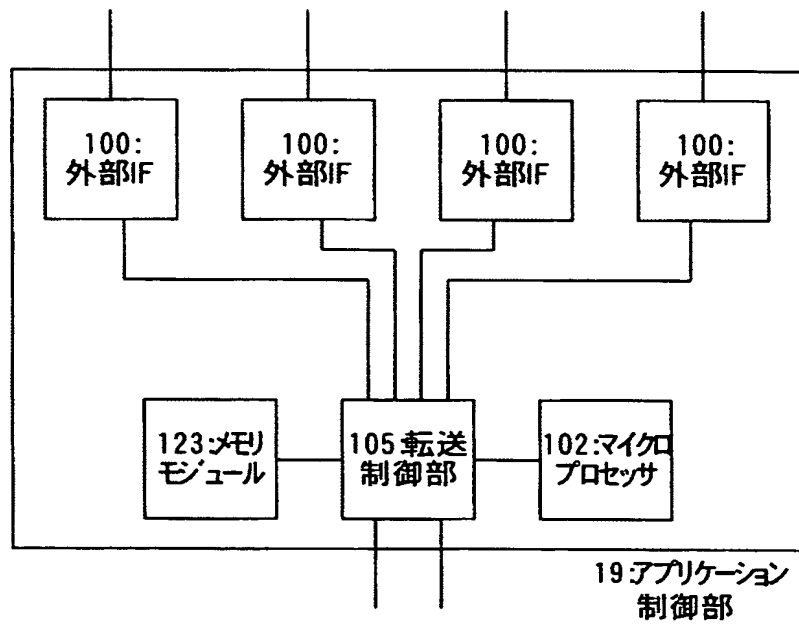
【図 12】

図12



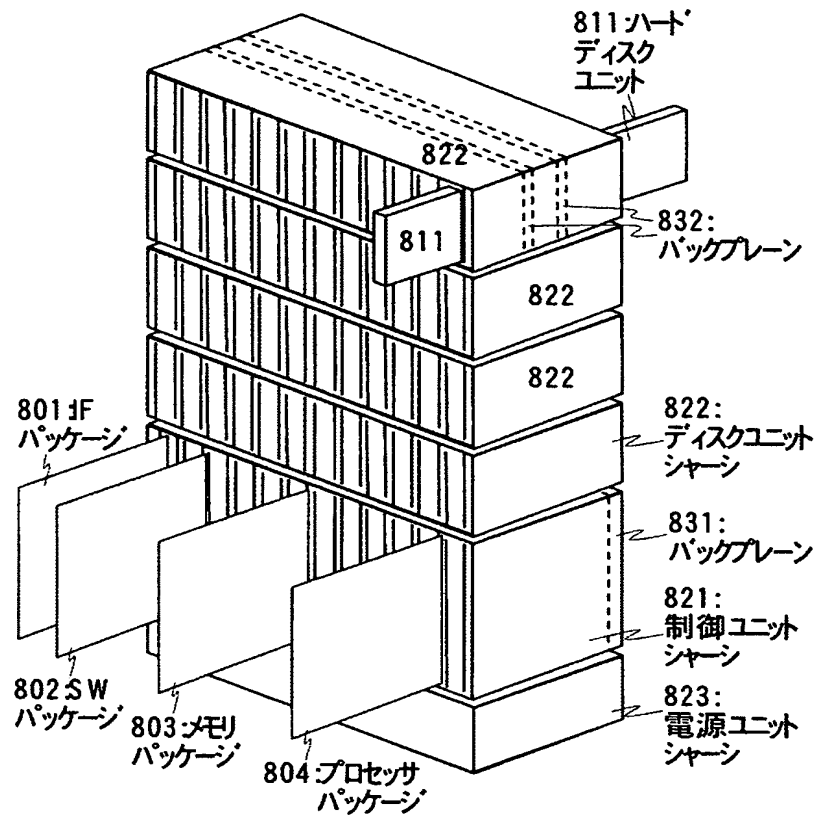
【図 13】

図13



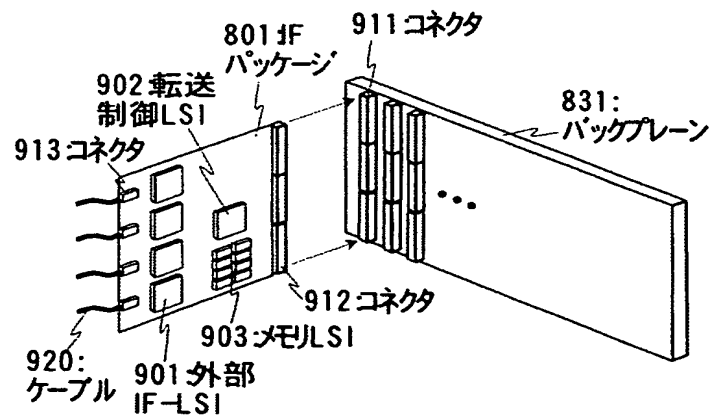
【図 14】

図 14



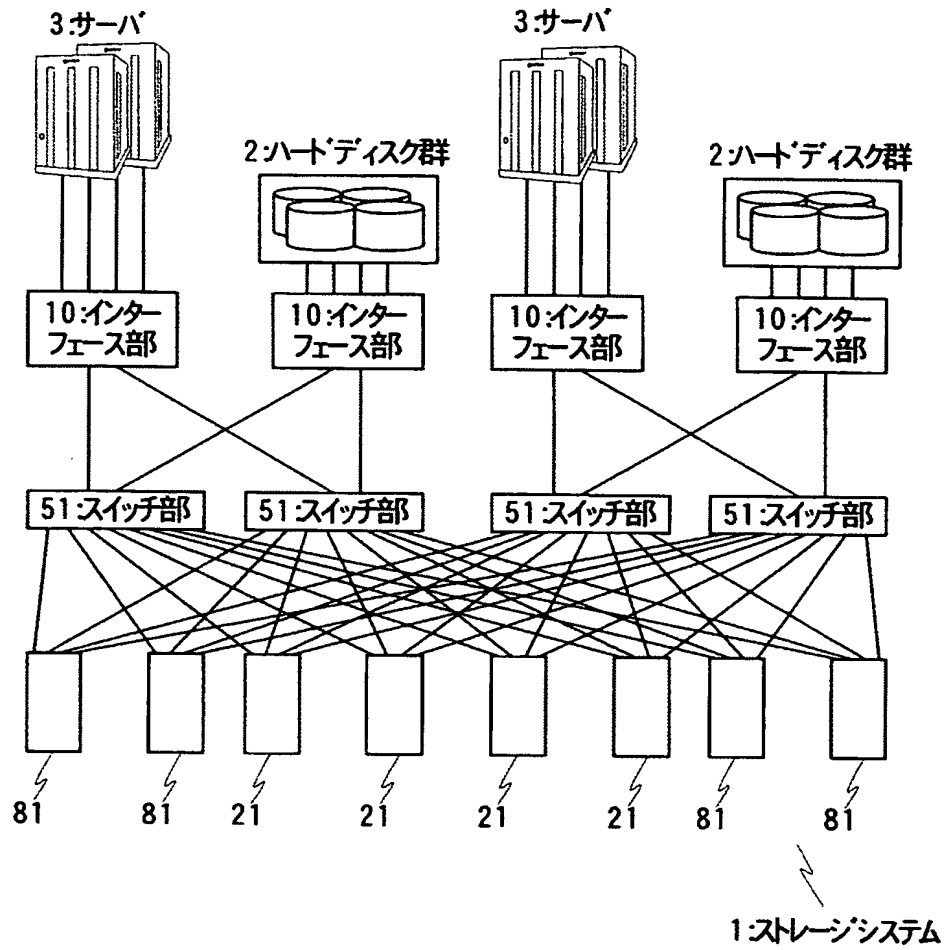
【図 15】

図 15



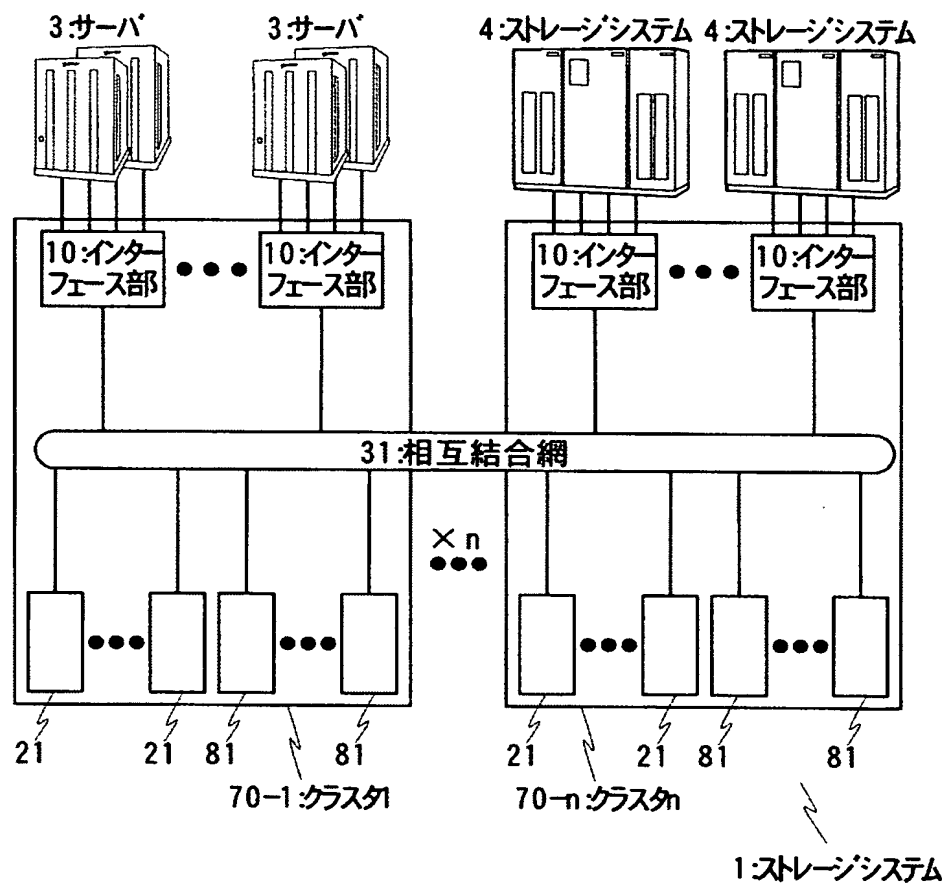
【図 16】

図 16



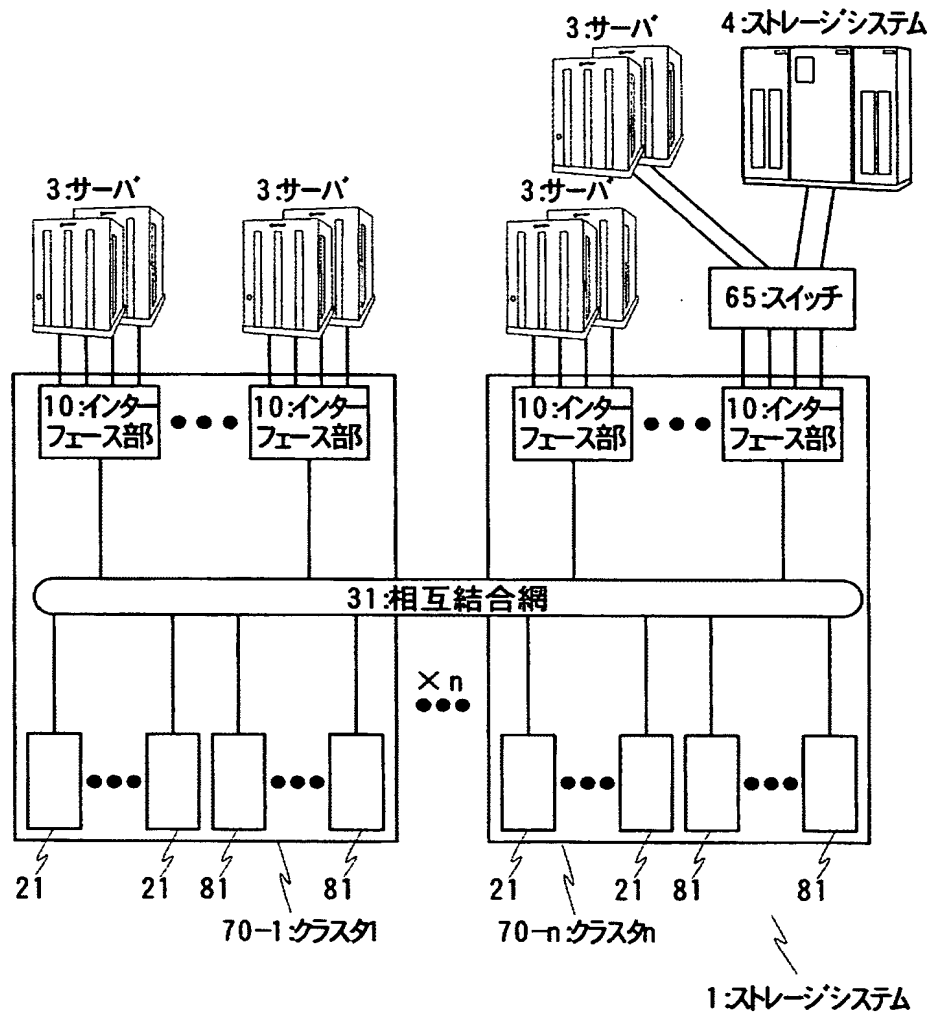
【図 17】

図 17



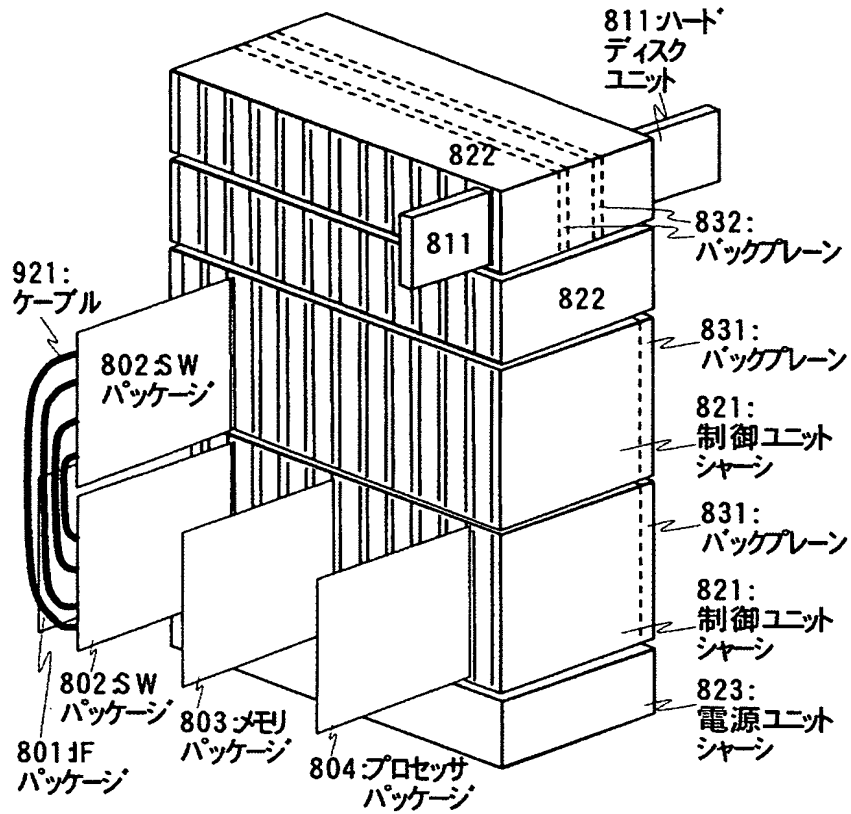
【図 18】

図18



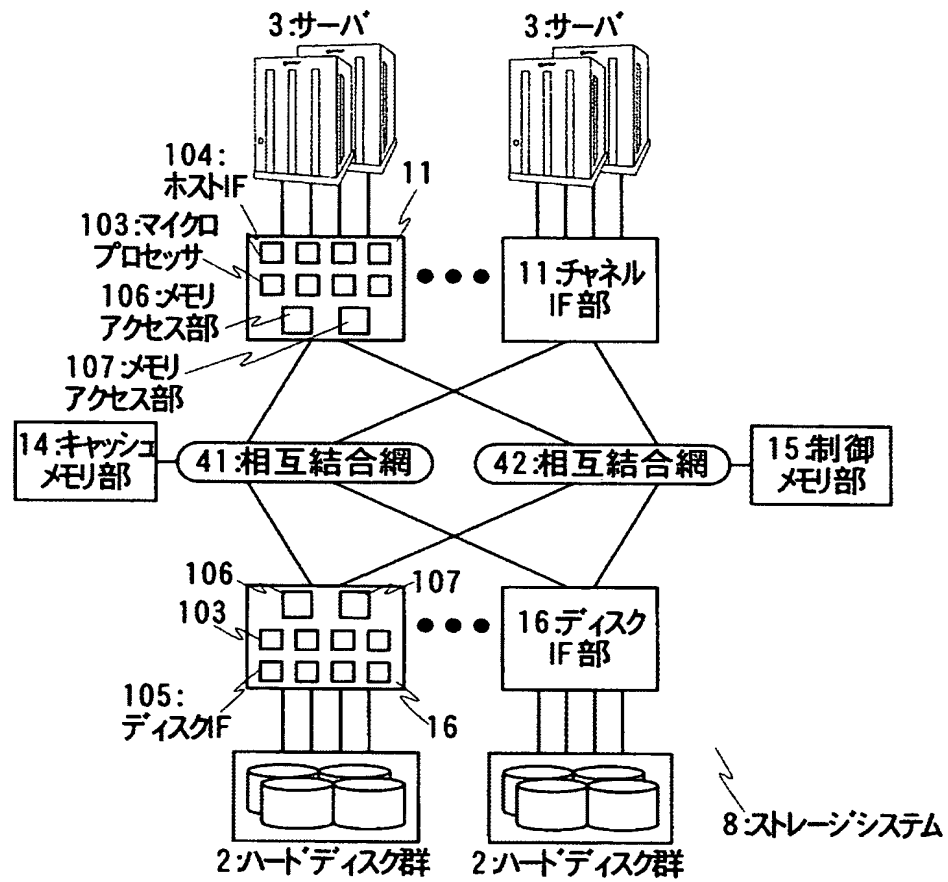
【図 19】

図19



【図 20】

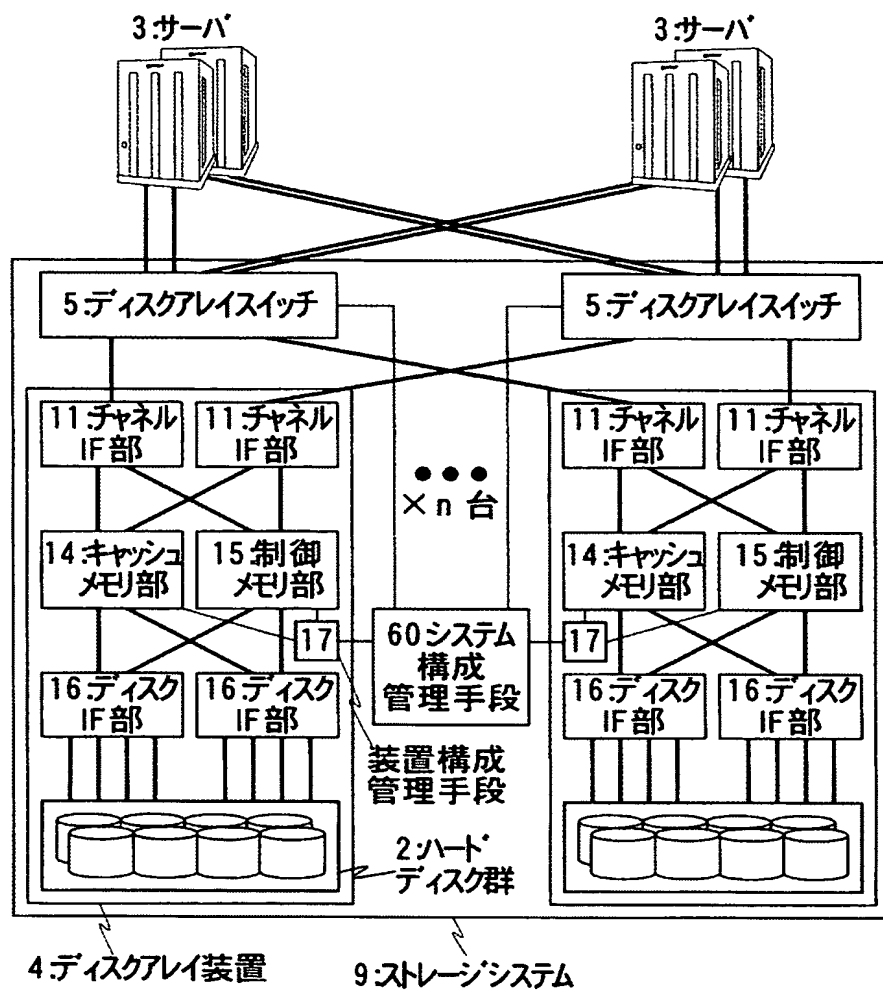
図20





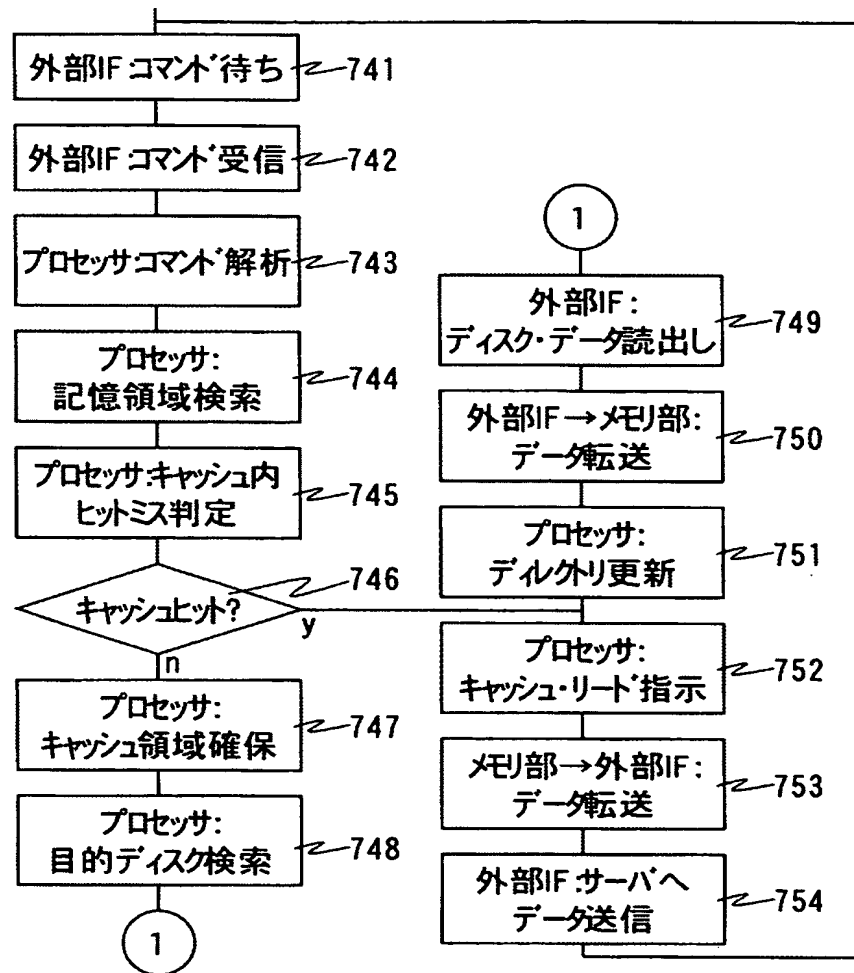
【図 21】

図21



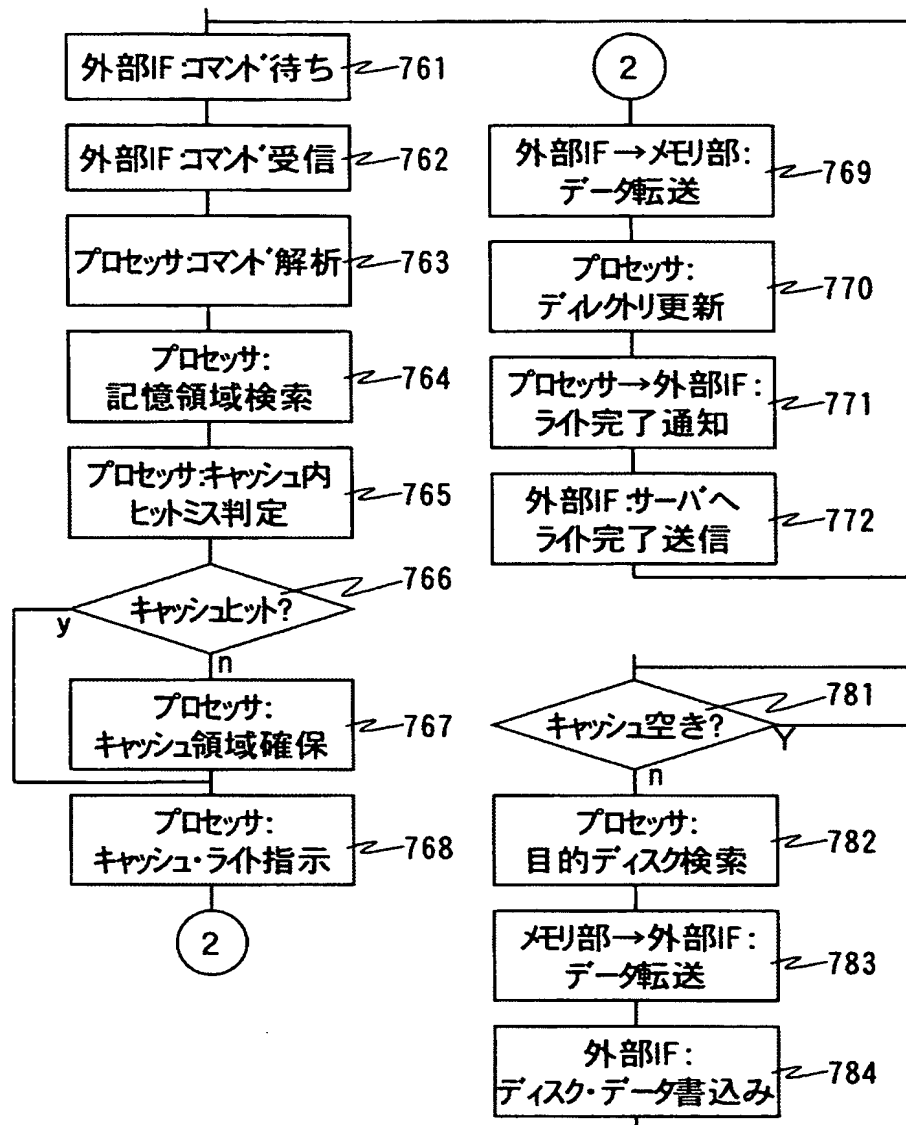
【図 22】

図22



【図 23】

図23



**【書類名】 要約書****【要約】****【課題】**

小規模～大規模な構成までユーザ要求に見合った価格／性能／接続性を実現可能なフレキシブルな構成のストレージシステムの提供。

**【解決手段】**

サーバ 3 あるいはハードディスク群 2 とのインターフェースを有するインターフェース部 1 0 と、サーバ 3 あるいはハードディスク群 2 との間でリード／ライトされるデータを格納するキャッシュメモリモジュール 1 2 6 とシステムの制御情報を格納する制御メモリモジュール 1 2 7 を有するメモリ部 2 1 と、サーバ 3 とハードディスク群 2 との間でのデータのリード／ライトを制御するマイクロプロセッサを有するプロセッサ部 8 1 と、相互結合網 3 1 からストレージシステムを構成し、インターフェース部 1 0、メモリ部 2 1 及びプロセッサ部 8 1 間を相互結合網 3 1 で互いに接続する構成とする。

**【選択図】 図 1**

認定・付加情報

特許出願の番号	特願 2 0 0 4 - 0 3 2 8 1 0
受付番号	5 0 4 0 0 2 1 2 2 4 4
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 6 年 2 月 1 2 日

< 認定情報・付加情報 >

【提出日】	平成16年 2月10日
-------	-------------

特願 2 0 0 4 - 0 3 2 8 1 0

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所